

Express Mail No. EL651822700US

FORM PTO-1390 (REV 10-95)		U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE		ATTORNEY DOCKET NUMBER	
<b>TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371</b>				15675P387	
INTERNATIONAL APPLICATION NO. PCT FR00/01734		INTERNATIONAL FILING DATE June 22, 2000		U.S. APPLICATION NO. (If known, see 37 CFR 1.5) <b>10/019550</b>	
TITLE OF INVENTION <b>A METHOD OF MANAGING THE DECODING AND PLAYBACK OF A SOUND SIGNAL IN AN</b>		PRIORITY DATE CLAIMED June 24, 1999			
APPLICANT(S) FOR DO/EO/US <b>David Deleam; Balazs Kovesi; Jean-Pierre Petit</b>					
Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:					
<ol style="list-style-type: none"> <li>1. <input checked="" type="checkbox"/> This is a <b>FIRST</b> submission of items concerning a filing under 35 U.S.C. 371.</li> <li>2. <input type="checkbox"/> This is a <b>SECOND</b> or <b>SUBSEQUENT</b> submission of items concerning a filing under 35 U.S.C. 371.</li> <li>3. <input type="checkbox"/> This express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b)) and PCT articles 22 and 39(1).</li> <li>4. <input checked="" type="checkbox"/> A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.</li> <li>5. <input checked="" type="checkbox"/> A copy of the International Application as filed (35 U.S.C. 371(c)(2)).           <ol style="list-style-type: none"> <li>a. <input checked="" type="checkbox"/> is transmitted herewith (required only if not transmitted by the International Bureau).</li> <li>b. <input type="checkbox"/> has been transmitted by the International Bureau.</li> <li>c. <input type="checkbox"/> is not required, as the application was filed in the United States Receiving Office (RO/US).</li> </ol> </li> <li>6. <input checked="" type="checkbox"/> A translation of the International Application into English (35 U.S.C. 371(c)(2)).</li> <li>7. <input type="checkbox"/> Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3)).           <ol style="list-style-type: none"> <li>a. <input type="checkbox"/> is transmitted herewith (required only if not transmitted by the International Bureau).</li> <li>b. <input type="checkbox"/> have been transmitted by the International Bureau.</li> <li>c. <input type="checkbox"/> have not been made; however, the time limit for making such amendments has <b>NOT</b> expired.</li> <li>d. <input type="checkbox"/> have not been made and will not be made.</li> </ol> </li> <li>8. <input type="checkbox"/> A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).</li> <li>9. <input type="checkbox"/> An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).</li> <li>10. <input type="checkbox"/> A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).</li> </ol>					
<b>Items 11. to 16. below concern document(s) or information included:</b>					
<ol style="list-style-type: none"> <li>11. <input type="checkbox"/> An Information Disclosure Statement under 37 CFR 1.97 and 1.98.</li> <li>12. <input type="checkbox"/> An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.</li> <li>13. <input type="checkbox"/> A <b>FIRST</b> preliminary amendment.  <input type="checkbox"/> A <b>SECOND</b> or <b>SUBSEQUENT</b> preliminary amendment.</li> <li>14. <input type="checkbox"/> A subsequent specification.</li> <li>15. <input type="checkbox"/> A change of power of attorney and/or address letter.</li> <li>16. <input checked="" type="checkbox"/> Other items or information:           <p style="margin-left: 20px;">priority request; formal drawings transmittal; 2 drawings sheets; preliminary examination report w/amended pages; request of filig; pct form 301 &amp; 304; verified English translation of amended pages; English translation of the report; copy of complete application w/amended pages (33 pages)</p> </li> </ol>					

U.S. APPLICATION NO. (if known, see 37 CFR 1.9) <b>10/019550</b>		INTERNATIONAL APPLICATION NO. <b>PCT FR00/01734</b>		ATTORNEY'S DOCKET NUMBER <b>15675P387</b>	
---	--	--	--	--	--

<b>17. <input checked="" type="checkbox"/> The following fees are submitted:</b> <b>BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5) ):</b> Neither international preliminary examination fee (37 CFR 1.482 nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by EPO or JPO ..... <b>\$1040.00</b> International preliminary examination fee (37CFR1.482)not paid to USPTO but International Search Report prepared by the EPO or JPO. .... <b>\$890.00</b> International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee paid to USPTO (37 CFR 1.445(a)(2)) ..... <b>\$740.00</b> International preliminary examination fee paid to USPTO (37 CFR 1.482) but all claims did not satisfy provisions of PCT Article 33(1)-(4) .... <b>\$710.00</b> International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(1)-(4) ..... <b>\$100.00</b>  <div style="text-align: right;"><b>ENTER APPROPRIATE BASIC FEE AMOUNT =</b></div> Surcharge of <b>\$130.00</b> for furnishing the oath or declaration later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(e)).				<b>CALCULATIONS FOR PTO USE ONLY</b>	
				\$ 890.00	
				\$	
CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total claims	15 - 20 =	0	X \$18.00	\$ 0.00	
Independent claims	2 - 3 =	0	X \$84.00	\$ 0.00	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ \$280.00	\$ 280.00	
<b>TOTAL OF ABOVE CALCULATIONS =</b>				\$ 1170.00	
Reduction of 1/2 for filing by small entity, if applicable. Verified Small Entity Statement must also be filed (Note 37 CFR 1.9, 1.27, 1.28).				\$	
<b>SUBTOTAL =</b>				\$ 1170.00	
Processing fee of <b>\$130.00</b> for furnishing the English translation later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(f)).				\$	
<b>TOTAL NATIONAL FEE =</b>				\$ 1170.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). <b>\$40.00</b> per property				\$	
<b>TOTAL FEES ENCLOSED =</b>				\$ 1170.00	
				\$ Amount to be: refunded	\$
				charged	\$

a. ☒ A check in the amount of \$ 1170.00 to cover the above fees is enclosed.

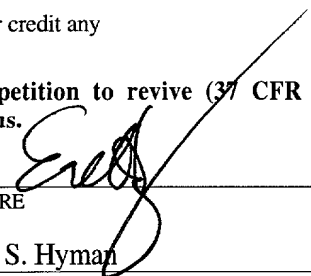
b. ☐ Please charge my Deposit Account No. \_\_\_\_\_ in the amount of \$ \_\_\_\_\_ to cover the above fees.  
A duplicate copy of this sheet is enclosed.

c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 022666. A duplicate copy of this sheet is enclosed.

**NOTE: Where an appropriate time limit under 37 CFR 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.**

SEND ALL CORRESPONDENCE TO:

Blakely, Sokoloff, Taylor & Zafman LLP  
12400 Wilshire Blvd. 7th Floor  
Los Angeles, CA 90025-1026

  
 SIGNATURE  
 Eric S. Hyman  
 NAME  
 30,139  
 REGISTRATION NUMBER

U.S. APPLICATION NO. (if known) <b>10/019550</b>	INTERNATIONAL APPLICATION NO. <b>PCT FR00/01734</b>	ATTORNEY'S DOCKET NUMBER <b>15675P387</b>
--	---	---

17. ☒ The following fees are submitted:**BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5) ):**

Neither international preliminary examination fee (37 CFR 1.482 nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by EPO or JPO ..... **\$1040.00**

International preliminary examination fee (37CFR1.482)not paid to USPTO but International Search Report prepared by the EPO or JPO. ... **\$890.00**

International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee paid to USPTO (37 CFR 1.445(a)(2)) ..... **\$740.00**

International preliminary examination fee paid to USPTO (37 CFR 1.482) but all claims did not satisfy provisions of PCT Article 33(1)-(4) .... **\$710.00**

International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(1)-(4) ..... **\$100.00**

**ENTER APPROPRIATE BASIC FEE AMOUNT =****CALCULATIONS FOR PTO USE ONLY**

\$ 890.00

Surcharge of **\$130.00** for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).

\$

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE
Total claims	15 - 20 =	0	X \$18.00
Independent claims	2 - 3 =	0	X \$84.00

\$ 0.00

\$ 0.00

MULTIPLE DEPENDENT CLAIM(S) (if applicable)

+ \$280.00

\$ 280.00

**TOTAL OF ABOVE CALCULATIONS =**

\$ 1170.00

Reduction of 1/2 for filing by small entity, if applicable. Verified Small Entity Statement must also be filed (Note 37 CFR 1.9, 1.27, 1.28).

\$

**SUBTOTAL =**

\$ 1170.00

Processing fee of **\$130.00** for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(f)).

\$

**TOTAL NATIONAL FEE =**

\$ 1170.00

Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). **\$40.00** per property

\$

**TOTAL FEES ENCLOSED =**

\$ 1170.00

\$ Amount to be: refunded

charged

\$

\$

- a. ☒ A check in the amount of \$ 1170.00 to cover the above fees is enclosed.
- b. ☐ Please charge my Deposit Account No. \_\_\_\_\_ in the amount of \$ \_\_\_\_\_ to cover the above fees. A duplicate copy of this sheet is enclosed.
- c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 022666. A duplicate copy of this sheet is enclosed.

**NOTE: Where an appropriate time limit under 37 CFR 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.**

SEND ALL CORRESPONDENCE TO:

Blakely, Sokoloff, Taylor & Zafman LLP  
12400 Wilshire Blvd. 7th Floor  
Los Angeles, CA 90025-1026

SIGNATURE

Eric S. Hyman

NAME

30,139

REGISTRATION NUMBER

Atty. Docket: 15675.P387  
Exp Mail EL651822700US

10/019550  
531 Rec'd PCT/PTC 21 DEC 2001

**UNITED STATES PATENT APPLICATION**

**FOR**

**A METHOD OF MANAGING THE DECODING AND PLAYBACK OF A SOUND  
SIGNAL IN AN ASYNCHRONOUS TRANSMISSION SYSTEM**

Inventor(s): DAVID DELEAM  
BALAZS KOVESI  
JEAN-PIERRE PETIT

Prepared by:

**BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN**  
12400 Wilshire Boulevard, 7<sup>th</sup> Floor  
Los Angeles, California 90025  
(310) 207-3800

10019550-032902

A METHOD OF MANAGING THE DECODING AND PLAYBACK OF A SOUND  
SIGNAL IN AN ASYNCHRONOUS TRANSMISSION SYSTEM

The present invention relates to a method of  
managing asynchrony in audio transmission.

5

**GENERAL DESCRIPTION OF THE FIELD OF THE INVENTION**

In general, the invention relates to transmission  
systems using low data rate speech encoders, in which the  
signals do not carry the reference clock of the source  
encoding system (the sampling frequency of the encoder).  
This applies, for example, to Internet protocol (IP) type  
transmissions or indeed to discontinuous transmissions,  
etc.

A general aim of the invention is to resolve the  
problems encountered by such systems in producing a  
continuous stream of decoded speech or sound.

Traditionally, telephone communications and sound  
channel networks have used analog transmission systems  
with frequency division multiplexing (primary groups,  
amplitude and frequency modulation). Under such  
conditions, the speech signal (or music signal; the term  
"speech" is used below throughout this document in  
generic manner) is converted into an electrical signal by  
a microphone and it is this analog signal which is  
filtered and modulated in order to be presented to a  
receiver which amplifies it prior to presenting it to a  
playback system (earphone, loudspeaker, etc.).

Over the last few years, digital transmission and  
switching techniques have been progressively replacing  
analog techniques. In pulse code modulation (PCM)  
systems, the speech signal is sampled and converted into  
a digital signal using an analog-to-digital converter  
(ADC) driven at a fixed sampling frequency derived from a  
master clock delivered by the network and also known to  
the receiver system. This applies to analog and digital  
subscriber connection units in telecommunications  
networks. The digital signal received by the destination

10019550 032900

(in the broad sense) is converted back into analog so that it can be heard by means of a digital-to-analog converter (DAC) driven by a clock at the same frequency as that used by the ADC of the source. Under such conditions, the entire system is entirely synchronous as generally applies to present-day switching and transmission systems. These can include data rate reduction systems (for example for a telephone signal, for converting from 64 kilobits per second (kbit/s) to 32 kbit/s or 16 kbit/s or 8 kbit/s). It is the network (or terminal systems as in the case of the integrated services digital network (ISDN) for example) which undertakes the operations of ADC, of encoding and decoding (where encoding and decoding are used in the context of reducing data rate), and of DAC. The clocks are always distributed and the system comprising ADC, speech encoding, transmission and switching, speech decoding, and finally DAC is fully isochronous. There are no losses or repeats of speech samples in the decoder.

The above-described synchronous transmission systems require the presence of a reference clock throughout the network. Transmission systems are now making greater and greater use (initially for data) of asynchronous and packet techniques (IP protocol, asynchronous transmission mode (ATM)). In numerous new situations, the decoder has no reference concerning the sampling frequency used by the encoder and it must be capable, using its own means, of reconstituting a decoding clock which attempts to track the reference of the encoder. The present invention is thus particularly advantageous in frame relay telephone systems, in ATM telephony, or in IP telephony. The technique described can easily be used in other fields of speech or sound transmission in which there exists no effective transmission of the clock reference from the encoder to the decoder.

## DESCRIPTION OF THE STATE OF THE ART

## The general problem

The general problem posed by transmission systems to which the invention applies is that of mitigating the fact that the speech or sound decoder has no clock reference associated with the source encoding.

In this respect, two circumstances can be distinguished: those corresponding to "weak" asynchrony and those corresponding to "strong" asynchrony.

*Weak asynchrony*

As an illustration, we consider the case of a transmission system comprising the following, as shown diagrammatically in Figure 1:

- an encoding source 1 comprising an analog-to-digital converter driven by a reference clock at frequency  $F_{ADC}$  equal to 8 kilohertz (kHz) (to provide numerical values in the worked examples below) and a speech encoder (of greater or lesser complexity and reducing the data rate for transmission to a greater or lesser extent);
- an asynchronous transmission system (represented by link 2) which conveys the information produced by the encoding source using its own transmission clock and its own protocols (for example the speech encoder could produce a data rate of 8 kbit/s and the transmission system could be constituted by an RS.232 type asynchronous link operating at 9600 bit/s); and
- a reception and decoding system 3 receiving the information conveyed over the asynchronous link (whose data rate must necessarily be a little greater than the raw encoding data rate, e.g. 9600 bit/s instead of 8000 bit/s) having the function of producing the signal after decoding (decompression) and applying the signal it produces to a digital-to-analog converter connected to a transducer such as a loudspeaker, a telephone handset, a

2065260" 0556Y00T

headset, or a sound card installed in a personal computer (PC).

It will be understood that since the reception and decoding system 3 has no clock reference, it must  
 5 implement a strategy in order to mitigate this asynchrony between the encoder and the decoder.

Whatever the encoding technique used or the type of transmission which does not directly convey a clock, or time markers within the transmitted frame, or indications  
 10 concerning transmission instants, the above-mentioned problem can be reduced (ignoring the speech encoder, the asynchronous transmission system, and the speech decoder) to a system comprising the following, as shown in Figure 2:

- 15 - an analog-to-digital converter 4 for converting speech signals or sound from analog to digital form at a sampling frequency set by a local oscillator;
- a digital-to-analog converter 5 for playing back the sound via a transducer suitable for the field of use  
 20 in question and operating at a sampling frequency given by a local oscillator which, a priori, is at the same frequency but which is never at exactly the same frequency for reasons of acceptable manufacturing cost (highly stable and very accurate frequency sources do  
 25 indeed exist, but they need to be temperature-compensated and they are unacceptably expensive for mass-produced industrial implementation); and
- a digital register 6 into which the analog converter 4 writes at its own sampling frequency ( $F_{ADC}$ ),  
 30 said register being read at the sampling frequency ( $F_{DAC}$ ) of the playback system by the digital-to-analog converter (DAC).

It will be understood that since the two clock frequencies ( $F_{ADC}$  and  $F_{DAC}$ ) are different, it is necessary  
 35 from time to time for the DAC to reread the same information twice over (if  $F_{DAC}$  is greater than  $F_{ADC}$ ) or on

10019550 032902



the contrary (where  $F_{DAC}$  is less than  $F_{ADC}$ ) to allow the ADC to overwrite information before the DAC can read it.

The oscillators that are commonly available in the trade are characterized by the accuracy with which they operate (within a certain temperature range).

Oscillators that are accurate to within 50 parts per million (ppm) are quite commonly available and are used to provide numerical values for the worked examples below showing how frequently samples are lost or repeated when the sampling frequency is 8 kHz (the reader can easily determine that at higher sampling frequencies samples are skipped or repeated at a frequency which is prorata the sampling frequency; the higher the sampling frequency the higher the frequency at which samples are skilled or repeated).

Under the least favorable conditions, an ADC is operating at  $8000 \times (1 + 50.e-6)$  in association with a DAC operating at  $8000 \times (1 - 50.e-6)$ . In this particular example, the skip period (period for samples being omitted in the DAC since  $F_{DAC}$  is less than  $F_{ADC}$ ) is easily calculated by counting the number of periods of the DAC (where the period is longer than that of the ADC) that produces a value equal to said period of the DAC when multiplied by the difference between the periods.

Writing the period of the DAC as  $P_{DAC}$  (in this case  $1/8000 \times (1 - 50.e-6)$ ) and  $P_{ADC}$  as the period of the ADC (in this case  $1/8000 \times (1 + 50.e-6)$ ) we obtain  $N \times (P_{DAC} - P_{ADC}) = P_{DAC}$ .  $N$  represents the number of individual operations that stem from the period difference. Writing  $50.e-6 = \epsilon$  and applying the simplifications that are common for small quantities, we obtain  $N = 1/(2\epsilon)$ . In this example, that immediately gives the skip period as being close to 1.25 seconds (s). If the accuracy of the local oscillators is improved (e.g. by going from 50.e-6 to 5.e-6) then the skip period will increase (in this case there will be one skip every 12.5 s).

In a complete transmission system including audio encoders operating on signal frames, this phenomenon of "slip" between two clocks will give rise to an absence of speech frames (no frame to be decoded in the time available for decoding) or to overabundance of frames (i.e. two frames for decoding instead of one in the available time). Taking the example of a speech encoder operating on 30 millisecond (ms) frames at 8 kHz, i.e. 240 samples, in each 30 ms time slot the receiver and more particularly the decoder expects to receive one frame for decoding in order to ensure that playback of the speech signal remains continuous. Unfortunately, if  $F_{ADC}$  is less than  $F_{DAC}$ , then on the above assumptions, there will be an absence of any frame of samples for decoding by the sound playback system once every  $240 \times 1.25 = 300$  s, and in the converse situation there will be two frames instead of one (i.e. a frame to be "eliminated") at the decoder once every 300 s. Under such circumstances, the awkward phenomenon of samples being skipped or repeated becomes very disagreeable since an entire block of the signal is skipped or repeated, and this needs to be managed appropriately.

### *Strong asynchrony*

Certain types of transmission amplify this problem of asynchrony due to the phenomenon of "slip" between clocks as explained above. This is what we refer to herein as "strong" asynchrony.

When transmission is imperfect, giving rise to samples or frames of samples being lost and also when transmission generates jitter on sample arrival times, where such jitter is associated neither with the sending clock nor with the receiving clock, but is associated with other mechanisms in the transmission system having their own clocks, then the receiver system can be confronted with an absence of several frames, or with an overabundance of several frames. This can apply, for

example, with IP type networks which suffer from the phenomenon of packets being lost and from the phenomenon of jitter introduced during packet routing. These phenomena disturb the continuity of the sound playback of the audio signal very strongly. When packets are lost or when jitter delays one or more packets, the playback system finds itself without any sample (or frame of samples) to apply to the DAC for the purpose of ensuring continuity in audio playback. Conversely, when jitter is strong, the playback system can find itself with far too many frames or samples to be sent simultaneously to the DAC. When jitter is strong, sound signal frame transmission can take place in the form of bursts, thus creating phenomena both of gaps and of overabundance amongst sample frames.

It will be observed that using speech encoders operating with a system of the voice activity detector/discontinuous transmission/comfort noise generation (VAD/DTX/CNG) type, a mechanism is also introduced that is similar to the loss of a packet since in the event of silence, the sender will cease to send frames of samples. Ceasing to send samples can be perceived at the receiver as being the same as the loss of a packet or as circumstances in which the ADC clock is faster than the DAC clock, which leads to holes in the signal at the receiver, as shown above.

"Strong" asynchrony thus differs from "weak" asynchrony by involving not only cyclical skips and/or repetitions, but also holes in the signal and/or overabundance of the signal in multiple and non-cyclic manner.

#### **Description of various existing methods**

Two main methods are presently known for mitigating the drawbacks due to the fact that the speech or sound decoder has no clock reference.

206240" 0556" 001  
The first consists merely in proceeding as described above in the paragraph describing "weak" asynchrony, i.e. by skipping or repeating samples. The decoding system produces samples at a rate that is more or less equal to that of the encoder and it presents them to the digital-to-analog converter at said rate (means for implementing the above reconstruction system are known to the person skilled in the art). In some cases, for example when "strong" asynchrony applies with transmission being in the form of frames, it is preferable when samples for playing back are missing to send null sample frames to the ADC rather than repeating a preceding frame. Furthermore, in the converse situation, when surplus samples are present, they are not eliminated directly, but a first-in-first-out (FIFO) register of some size can be used to absorb jitter to some extent. If the FIFO register becomes too full, then that triggers partial or complete emptying of the FIFO, thereby giving rise to new skips in sound playback.

20 The second method, which is more complicated and provides better performance, requires a loop to be implemented to recover a hardware clock which is servo-controlled by the filling level of a buffer memory for the signal to be decoded (or to be transmitted as in the ATM adaptation layer number 1 (AAL1) for example). That method of servo-control attempts to use the clock recovery loop to recover the sampling frequency of the source. The filling level of the receive buffer produces a control signal for servo-controlling a digital or analog phase-locked loop (PLL).

30 The first method is extremely simple to implement but suffers from a major defect associated with the quality of the sound reproduced. A skip or elimination once every 1.25 seconds can be very disagreeable to listen to, and this can occur with "weak" asynchrony associated with correction at sample level. Similarly, for a system operating with frames of samples, the

inserted repetitions or blanks, and the discontinuities in the signal due to frames being eliminated amplify loss of quality which becomes highly perceptible and very disturbing for the listener.

5 Furthermore, the use of a FIFO memory runs the risk of establishing a considerable delay in transmission and that also harms the overall quality of a call.

10 The second method is much more complex to implement and requires a clock servo-control mechanism, and thus requires special hardware. However, it provides partial synchronization and therefore avoids problems associated with managing asynchrony. Nevertheless, that method adapts poorly to discontinuous transmission systems, to systems involving last frames, or to systems with high  
15 levels of jitter. Under such circumstances, synchronization information is no longer available. Furthermore, that method cannot be envisaged on terminal platforms where clock servo-control is not possible, as is the case in particular with PC type terminals, for  
20 example, where the system used for playing back sound is a sound card.

Devices are already known from document WO/99 17 584 for implementing a method in accordance with the preamble of claim 1, the devices having only one buffer memory.

25 Document US-A 4 703 477 facilitates reading voice data by implementing a method of putting frames relating to the same voice data end-to-end.

#### SUMMARY OF THE INVENTION

30 A general object of the invention is to propose a solution to the problems associated with continuity in the playback of a speech signal in the presence of asynchronous transmission, and to do so by taking action at receiver level, i.e. at the end of the transmission  
35 system.

To this end, the invention provides a method of managing the decoding and playback of a sound signal in

CORRECTED SHEET

an asynchronous transmission system, in which any overabundance of filling of a first buffer memory and/or of a second buffer memory situated at the inlet or at the outlet of a decoding block is detected by comparing the filling level with at least one threshold, the method  
 5 being characterized in that, depending on the value of the filling level:

- voice activity detection is implemented and frames considered by said detection as being non-active are  
 10 eliminated; and

- concatenation processing is implemented on two successive frames to compact them into a pseudo-frame of length less than or equal to one frame, the length reduction ratio of the pseudo-frame relative to the  
 15 length of the two frames being greater than or equal to two.

Such a method is simple to implement and provides a guarantee of quality by avoiding excessive increase in transmission delay and by managing holes in the speech  
 20 signal effectively. Furthermore, it does not imply any specific hardware servo-control circuit, and can therefore be quickly adapted to different asynchronous networks, terminals, and platforms.

The method is advantageously associated with the  
 25 various characteristics below taken singly or in any technically feasible combination:

- voice activity detection is implemented and frames considered by said detection as being not active are eliminated whenever the filling level lies between a  
 30 first threshold and a second threshold, and in that concatenation processing is implemented on two successive frames whenever the filling level lies between a second threshold and a third threshold;

- the first and second thresholds are the same;  
 35 - detection is performed at the inlet or the outlet of a decoding block having a first buffer memory at its inlet and/or its outlet to determine whether any frame is

CORRECTED SHEET

- missing or erroneous or whether any samples to be played back are absent, and a fake frame is generated to ensure continuity in the audio playback on detecting such a missing or erroneous frame, or on detecting such an absence of samples for playback;
- 5

- when the decoding block implements its decoding processing in cyclical manner relative to the content of the first buffer memory, detection of any missing or erroneous frame or of any absence of samples to play back

206220" 05561001

CORRECTED SHEET

This translation of an amended page covers the amendments made in the original. However, the page breaks match the translation, so that this page is also a replacement page that fits in with the remainder of the translation.

is implemented at the same cyclical frequency, said detection taking place far enough in advance relative to the decoding process to make it possible to generate a fake frame in good time;

- 5           - a fake frame is not generated when a missing or erroneous frame is detected for a frame on which an absence of samples has already been detected;
- for a system of the type which can voluntarily stop sending frames, the type of the previously-generated
- 10       frame is stored from one frame to the next, and this information is used to determine whether to generate fake frames or to generate frames of silence;
- in processing for concatenating two successive frames, the samples are weighted in such a manner as to
- 15       give more importance to the first samples of the first frame and to the last samples of the second frame;
- the threshold(s) is/are adaptive; and
- a threshold is adapted as a function of the length of time passed with a filling level above a given
- 20       threshold.

The invention also provides a device for playing back a speech signal, the device comprising a first buffer memory receiving coded frames, means implementing decoding processing on the frames stored in said first

25       buffer memory, a second buffer memory receiving decoded frames output by the decoding means, and sound playback means receiving the frames output by the second buffer memory, the device being characterized in that it further comprises means for implementing the above-specified

30       method.

As will be understood on reading the following description, these means are essentially computer means.

#### DESCRIPTION OF THE FIGURES

35       Other characteristics and advantages of the invention appear further from the following description which is purely illustrative and non-limiting and which

10019550.032903



should be read with reference to the accompanying figures, in which:

- Figure 1 is a block diagram of an asynchronous transmission system;
- 5       - Figure 2 is a diagram showing a model of such a transmission system;
- Figure 3 is a diagram of a receiver device; and
- Figure 4 shows the signals obtained by implementing concatenation processing as proposed by the invention.

#### DETAILED DESCRIPTION OF ONE OR MORE EMBODIMENTS

The method proposed by the invention for managing asynchrony in transmission implements two kinds of processing corresponding to handling the two phenomena described above, namely the lack of samples and surplus samples.

##### 1. Description of the sound playback system in a conventional transmission application

As shown in Figure 3, the playback system for a speech signal comprises three elements:

- A block 10 waiting to receive samples or frames of code coming from the network. The block 10 contains a FIFO type memory 11 or circular type buffer memory (referred to as "FIFO 1" in the description below) enabling frames to be stored on a temporary basis prior to being decoded.

- A decoding block 12 which takes the frames coming from the block 10, decodes them, and stores them in turn in a FIFO memory 13 (referred to below as "FIFO 2").

- A playback block 14 which takes the decoded sample frames and applies them to any kind of sound playback system 15.

Depending on the terminals and the way the system is organized, the clock frequency used for sound playback (i.e. the digital-to-analog converter frequency  $F_{DAC}$ ) is

not necessarily directly associated with all of the blocks. Since the block 14 is directly associated with the playback system, it is directly associated with the frequency  $F_{DAC}$ . However the other blocks can be

5 associated instead with the rate at which frames arrive from the network rather than with the frequency  $F_{DAC}$ . Taking the example of a terminal provided with a multitasking system, and in which each block is performed by a specific task, the tasks 10 and 12 can thus be  
10 associated with frame reception. The task 10 waits for a frame from a network, which frame is then decoded by the task 12 and placed in the memory FIFO 2.

Meanwhile the task 14 clocked at  $F_{DAC}$  takes samples from the memory FIFO 2 and delivers them continuously to  
15 the sound playback system.

It can thus be seen that regardless of whether the asynchrony is "strong" or "weak", it is the way in which the memory FIFO 2 is managed that requires particular attention. Similarly, if the task 12 were strongly  
20 associated with the task 14, then particular attention would be required by the memory FIFO 1.

The mechanism constituting an implementation of the invention is described below in application to managing the memory FIFO 2, but the description includes  
25 explanations about how to transpose it with certain adaptations to managing the memory FIFO 1.

## 2. Absence of samples

In order to continue playing back sound in the  
30 absence of samples, both potential causes of samples for playback being absent are treated. The first cause corresponds to information contained in lost packets, while the second cause corresponds to the absence of any samples to play back (e.g. FIFO 2 empty) even though it  
35 is still necessary to keep on sending samples to the sound playback system.

## 2.1 Loss of frames or erroneous frames

The processing applied to lost frames or to erroneous frames requires a transmission system to be available that gives access to information about frames being lost and about erroneous frames being received. This is often the case in transmission systems.

For example, in IP networks, it is possible to use the marking of packets coming from the real time transfer (RTP) layer, which marking gives the exact number of samples lost between two packets of audio code being received. This information about loss of frames, or in the case of IP about loss of packets (each containing one or more speech frames) generally becomes available only once the packet following the lost packet(s) is itself received.

It is not necessarily advantageous to take action, while one or more valid frames can be decoded. With new generation speech encoders (CELP encoders, transform encoders, ...), in order to ensure that the quality of sound playback is maintained, it is often necessary to ensure a degree of synchronism between the encoder and the decoder. The loss of this encoder/decoder synchronism can be compensated by using frame loss correction algorithms associated with the speech encoder used. By way of example, these algorithms are provided in the standards for certain speech encoders (e.g. International Telecommunications Union (ITU) standard G.723.1). When using simpler encoders, such a mechanism is not always necessary.

When a large number of frames have been lost, the number of "fake" sample frames that need to be generated in order to pack out the memory FIFO 2 can be limited. The purpose of processing fake frame generation is to fill holes in such a manner as to ensure signal continuity while also smoothing the internal variables of the decoder so as to avoid excessive divergence on decoding the first valid frame following the invalid or

10019550 032900

lost frames, thereby avoiding any audible discontinuity. After a few frames have been generated, it can be assumed that the variables have been smoothed, and thus that the generation of such fake frames can be limited to a small number of frames (e.g. four to six) whenever a large number of frames has been lost.

It will thus be understood that processing is servo-controlled in this way relative to information about lost frames.

Similar processing is implemented on the basis of information about invalid frames. This information is forwarded to the decoder by the network portion of the receiver and it arrives soon enough to enable a frame correction algorithm to be implemented which, by taking account of such a non-valid frame, makes it possible to ensure continuity in the signal, and thus to avoid having another cause for samples being absent in the memory FIFO 2.

To sum up, this first process corresponds to managing information of the type "n frames lost" or "invalid frame received" coming from the network layer of the receiver. This management is characterized by implementing an algorithm for correcting frame losses (also referred to in this document as an algorithm for generating "fake" frames). This first process therefore acts at decoding task level and feeds the memory FIFO 2.

## 2.2 Absence of samples to be played back

This second process is associated with the clock coming from the task 14, i.e. with the clock at the frequency  $F_{DAC}$ . As mentioned above, the memory FIFO 2 (or FIFO 1 if the task 12 is included in the task 14) can become empty of samples even though it is still necessary to supply samples to the sound playback system. It is then necessary to supply the playback system with samples and if possible to avoid playing back zeros (since this degrades the sound signal very greatly).

This second process can be thought of as a feedback loop on frame decoding. This loop causes the algorithm for correcting frame losses to be called and as a result it needs to be activated soon enough to enable the

5 algorithm to be executed and to enable samples to be sent to the sound playback system. Depending on the platform, this feedback can be called in different ways.

This loop can be implemented in two ways which are described below.

10 For a single-task receiver (e.g. a digital signal processor (DSP) without any real time operating system (RTOS)), the audio decoder portion is tied completely to the DAC clock ( $F_{DAC}$ ) and is therefore permanently waiting for a frame to be decoded in cyclical manner. For

15 example, with a speech encoder using 30 ms frames, waiting loops are built up in periods that are multiples of 30 ms.

Thus, for a 30 ms loop, the decoder will, every 30 ms, be expecting a frame for decoding to be placed in the memory FIFO 1 (which can correspond merely to a frame passing from the network layer to the task 12). On

20 arrival of the frame, it is decoded and placed in the memory FIFO 2 for sending to the DAC. The feedback processing is implemented whenever it is observed that

25 there is no frame for decoding in the memory FIFO 1 at the time given by:

$$T = T_o + 30 \text{ ms} - T_c$$

where:

$T_o$  = the start time of the 30 ms wait loop; and

30  $T_c$  = the time required for executing the algorithm for generating fake frames with a safety margin corresponding to interrupts and/or other auxiliary processing that might take place before the end of the loop.

35 Processing is thus implemented with a latency time deadline of  $T_b$  (loop time) -  $T_c$  (computation time + margin).

With a multitasking receiver (e.g. a PC terminal), time is not managed with such precision and the processing implemented must therefore be somewhat different. (Note: this processing nevertheless remains quite close to the preceding process since it too seeks to take account of the time  $T_c$ .)

Under such circumstances, the only waiting loops available are often those associated with events, e.g. the fact that packets have been received by the network, or the fact that buffer memory  $n$  (containing one or more sample frames) sent previously to the sound playback system has been read by the DAC and is therefore again available for sending samples to the DAC.

Depending on the structure of the system and on whether or not it is necessary to respond quickly to an event, it is possible to wait for a certain length of time before filling said buffer memory prior to forwarding to the DAC. Such a latency time is selected in such a manner as to leave enough time for the algorithm for generating "fake" frames to execute, if necessary.

Then, possibly after said latency time has elapsed, the process verifies that sufficient samples are present in FIFO 2 (note: this could apply to FIFO 1 if the management takes place at its level), and if not it requests an appropriate number of fake frames to be generated in order to fill buffer memory  $n$ .

When the system is such that it is necessary to fill buffer memory  $n$  "immediately", then monitoring the availability of samples and possibly calling for the "fake" frames generation processing are implemented directly after each delivery to the DAC from the buffer memory so that the generated samples are already in the memory FIFO 2 when the event "buffer memory  $n$  available" occurs.

Thus, whatever the receiver, the process observes the absence of samples to be sent to the sound playback

10019550 032900

system by implementing a check on the content of the buffer memory FIFO 2 (or FIFO 1 depending on how the sound playback system is managed) and activates the algorithm for generating "fake" frames in order to generate the missing samples.

It will be understood that the second process responds firstly to the problem of "slip" between clocks, and more precisely to the circumstance in which the received clock ( $F_{DAC}$ ) is faster than the send clock ( $F_{ADC}$ ). It also applies to the phenomenon of frames being lost since this can lead to there being an absence of samples to send to the DAC even before frame loss has been detected, since such detection occurs only on receiving the frame following the loss.

In order to combine the actions of the first and second processes, the first process is prevented from generating "fake" frames on detection of frame loss whenever the corresponding frames have just been generated by the second process.

For this purpose, use is made of flags and also of counters determining the number of samples that have been generated by the second process.

### 2.3 Specific actions for speech encoders using VAD/DTX/CNG services

Encoders using a VAD/DTX/CNG system can voluntarily stop sending frames; under such circumstances, the absence of samples must not be considered exactly as a loss of frames, but rather as a period of silence. The only way of determining whether the frame to be generated must be silence or should correspond to a lost frame is to know the type of the previously-generated frame (i.e. signal frame or frame corresponding to a lost frame, or a noise update frame (SID), or a frame of silence (NOT)). For this purpose, the type of the generated frame is stored, and while frames are being generated to compensate for an absent frame or a lost frame, it is

decided whether fake frames should be generated using the algorithm for correcting frame losses (as applies when the preceding frame was of the FSF type), or whether frames of silence should be generated by activating the decoder appropriately (as applies when the preceding frame was of the SID or the NOT type).

### 3. Overabundance of samples to be played back

In order to deal with an overabundance of samples to be played back, processing is implemented to empty out frames, eliminating certain frames in full or in part prior to their possibly being taken into account by the sound playback system.

This processing enables frames to be stored in buffer memories until certain thresholds trigger actions for limiting the amount of frames in memory and for limiting any corresponding increase in delay across the communications system. This limited storage makes it possible to accommodate jitter phenomena on receiving frames in bursts and also slip between clocks, while nevertheless limiting transmission delay.

#### 3.1 Emptying out processing

Any accumulation of frames is initially detectable in the memory FIFO 1, and is subsequently transferred to the memory FIFO 2.

The proposed method manages information concerning the filling level of a reference buffer memory, i.e. FIFO 1 or FIFO 2 depending on how the tasks 10, 12, and 14 are organized in the receiver (as explained above). If the tasks 12 and 14 are associated with each other, then the filling level information used by the method relates to the memory FIFO 1 which acts as a buffer between the network and the sound playback system. Similarly, if the tasks 10 and 12 are associated, then it is the memory FIFO 2 which acts as a buffer and it is



therefore its filling level which is taken into consideration by the management process.

The process is explained below for the second context. The first is merely an immediate transposition thereof.

5 In order to maintain synchronization as closely as possible between the encoder and the decoder, and thus maintain optimum sound playback, all of the frames coming from the network are decoded. The process then decides  
10 on what action to take on the decoded frame as a function of information concerning filling level. This action is described in greater detail below. To activate the processing, filling level thresholds are used. These thresholds define filling alarm levels for the FIFO  
15 memory. In order to take action that is as inaudible as possible (i.e. in order to limit quality degradation) two levels of action are selected. A first level (alarm level 1) corresponds to the filling level being excessive but not critical (far from the maximum acceptable filling level), while a second level (alarm level 2) corresponds  
20 to it being mandatory to take action on each frame (this level is quite close to the maximum acceptable level). A third or "emergency" level (alarm level 3) is also defined in order to avoid memory overflows or other  
25 problems. This level corresponds to filling being very close to the maximum acceptable level. Alarm level 3 should never be reached if the actions taken at the two preceding threshold levels are properly performed and if the thresholds are properly defined.

30 Each time decoding is performed, the information concerning filling level is compared with the thresholds in order to determine the state of the FIFO (in an alarm state or not), and, where appropriate, the level of the alarm.

35 If the state obtained is not an alarm state, then no action is undertaken and the decoded frame is stored in FIFO 2.

10019550.032900

In the first alarm state, it is considered that at least 50% of the signal coming from a conversation is not useful and therefore at this alarm level, all frames presenting very little information are eliminated. For this purpose, it is possible to implement simple VAD which monitors all frames of samples after they have been decoded to decide whether or not they should be written into FIFO 2. The process can also make decisions based on information taken directly from the code frame concerning the importance or otherwise of the information contained in the frame. In this alarm state, any frame that is considered as containing nothing but noise is simply not stored in FIFO 2 for future sound playback.

In the second alarm state (critical level), it is necessary to take action on each frame to curb growth in the filling level of the memory FIFO 2 very aggressively. At this level, the preceding processing (i.e. the processing implemented for alarm level 1) remains active. However it is now also necessary to shorten pairs of consecutive frames down to the length of one frame or shorter. A decision is therefore taken on the basis of two non-"silent" sample frames (given that any frame that is "silent" is merely not written to FIFO 2 as a result of alarm state 1 being already included in alarm state 2). Action on two consecutive frames is therefore undertaken only when a frame is detected as being non-"silent". The frame is initially stored, and then if the next frame is "silent", then it is only the first frame that is written into FIFO 2.

When both frames contain important information, it becomes necessary to replace them by a single frame while minimizing loss of information and degradation of quality. It is the replacement frame that is stored in FIFO 2. Any effective solution capable of performing this task can be used and activated under such conditions (i.e. second alarm state and two non-"silent" frames).

Two examples of algorithms for performing this task are described below.

In a first algorithmic solution, the two contiguous frames are replaced by a single frame in which each  
 5 coefficient  $x_j$  (where  $j$  lies in the range 0 to  $N-1$  and where  $N$  is the number of samples per frame) is given the value  $(x_i + x_{i+1})/2$  (where  $i$  lies in the range 0 to  $2N-1$ , with the coefficients  $x_i$  coming from both original frames). This solution amounts to performing a kind of  
 10 smoothed undersampling. The frequency of the played-back signal is thus doubled for this frame. Nevertheless, the inventors have found that providing alarm state 2 does not occur very frequently, this solution suffices to maintain the quality of sound playback.

15 In a second solution, signal amplitude is detected to enable the two frames to be compacted into a pseudo-frame of length shorter than or equal to that of one frame. The number of samples contained in the pseudo-frame is determined by the fundamental frequency or  
 20 "pitch" information, but in all events it is shorter than or equal to the length of a normal frame, while nevertheless being close to said frame length. The algorithm used ensures continuity of the played-back signal without any audible hole, and without any  
 25 frequency doubling, while nevertheless dividing the amount of storage required for the signal by a factor that is greater than or equal to 2. This is explained in greater detail in paragraph 3.4 below. Furthermore, this also minimizes the loss of sound information since less  
 30 than 50% of the information is in fact eliminated.

It should be observed that when the receiver implements its processing on the basis of analyzing  
 FIFO 1, with the decoder being directly associated with the sound playback system, it is necessary to generate a  
 35 number of samples that is sufficient, i.e. in the present case a number that ensures at least one frame of samples is made available in FIFO 2. The frame concatenation

algorithm is then calibrated to ensure that it always generates a minimum number of samples, but at least one frame. Another solution would consist in activating the algorithm several times over instead of only once when that is necessary to obtain a sufficient number of samples.

In the third alarm level, which is normally never reached, no frames are stored in FIFO 2. In a variant, the system can also decide to clear out a fraction of the buffer memory suddenly (this can apply where it is management of FIFO 1 that is activated).

It should also be observed that depending on the network and on the types of problems it generates in terms of asynchrony, it is possible to decide whether or not to activate particular alarm levels. For example, when asynchrony is "weak" then alarm levels 1 and 2 can be combined, and the simple solution of replacing two frames by a single frame can then constitute the only active process.

### 3.2 Alarm thresholds

There follows a more detailed description of the alarm thresholds and how they are managed.

As explained above, the reference memory is said to be in alarm state 1 when its filling level is above threshold 1; this state remains active until its filling level drops below threshold 0. State 1 therefore operates with hysteresis.

The memory is said to be in alarm state 2 if the filling level exceeds threshold 2 and to be in alarm state 3 if the filling level exceeds threshold 3. It is possible to envisage managing these alarm states with hysteresis as well.

Thresholds 0, 1, and 2 are adaptive. Threshold 3 is directly associated with the maximum acceptable size and it is fixed. These thresholds need to be adaptable in order to accommodate different call contexts and real

time fluctuations during a call. It is appropriate to allow a greater amount of delay when the call is suffering a large amount of jitter (where delaying playback remains the best way of ensuring acceptable quality in the presence of jitter). In a high-jitter context, it is therefore appropriate for the thresholds 0, 1, and 2 to be at quite high levels.

To facilitate processing, the positions of the thresholds can correspond to integer numbers of frame sizes as exchanged between the various tasks of the receiver. This frame size is written  $T_t$ .

By way of example, the initial values of these thresholds can be as follows:

Threshold 0:  $5 \times T_t$   
 Threshold 1:  $8 \times T_t$   
 Threshold 2:  $12 \times T_t$   
 Threshold 3:  $24 \times T_t$  (fixed value)

The thresholds 0, 1, and 2 can be adapted together in steps of size  $T_t$ . Extreme acceptable values can lie in the range -1 to +8, for example.

Thus, threshold 1 can take on values  $7 \times$ ,  $8 \times$ ,  $9 \times$ ,  $10 \times$ , ...,  $16 \times T_t$ . Threshold adaptation proper is performed on the basis of an adaptation criterion which is the length of time spent in the alarm state. For this purpose, an alarm state percentage is evaluated about once every  $N$  seconds (e.g.  $N = 10$ ). When this percentage exceeds a given threshold (5%), the alarm state thresholds are increased; otherwise, when these percentages are below a given minimum threshold (0.5%) the alarm threshold are decreased. To ensure that the system does not oscillate excessively due to its thresholds being adapted too frequently, hysteresis is applied to adaptation decision making. The thresholds are actually increased by one step, only in the presence of two increase options that are consecutive, and they are decreased by one step, only in the presence of three decrease options that are consecutive. As a result, the length of time between two

10019550-032900

threshold increments is at least  $2N$  seconds and the length of time between two threshold decrements is at least  $3N$  seconds. The procedure for increasing thresholds can be accelerated if a large percentage of frames are in an alarm state. One accelerating procedure consists in increasing the thresholds directly, for example whenever the alarm percentage exceeds 50%.

Naturally, the threshold values given for the alarm thresholds are provided purely by way of indication.

### 3.3 Interaction with the first process

The first process is the process which causes "fake" frames to be generated when frames are lost or erroneous. When the system is in an alarm state (overabundance of frames), there is no need to generate "fake" frames which would merely aggravate the phenomenon of overabundance. Nevertheless, in order to maintain high quality sound playback it is important to maintain encoder/decoder synchronization by informing the decoder whenever a frame has been lost (e.g. by launching the generation of one or two fake frames, but no more). The third process will act in the alarm state on the first process so as to curb very strongly the generation of "fake" frames.

### 3.4 Frame concatenation

The purpose of the concatenation process is to shorten the duration of a digital audio signal containing speech or music while introducing as little audible degradation as possible. Since the sampling frequency is given and fixed, it is the number of samples sent to the sound playback apparatus that is decreased. One obvious solution for shortening a sequence of  $N$  samples is to remove  $M$  regularly spaced apart samples from the sequence in question. This causes the fundamental frequency to increase and that can be unpleasant for the listener, particularly when the ratio  $M/N$  is large. Furthermore, there is a danger of no longer complying with the

10019550.032900

sampling theorem. The process described below makes it possible to shorten an audio sequence without modifying its fundamental frequency and without giving rise to audible degradation due to signal discontinuity. This process is based on detecting the value of the pitch period. The number of samples eliminated by this algorithm cannot be selected freely, since it is a multiple of the pitch value  $P$ . Nevertheless, it is possible to define a minimum number of samples to be eliminated  $N_{emin}$  which must satisfy the relationship  $N_{emin} \leq N/2$ . Since the purpose is to eliminate at least 50% of the samples in the context of the device for managing asynchrony in an audio transmission, it is advantageous to set  $N_{emin} = N/2$ . It is also assumed that the maximum value of the pitch  $P$  is less than the length  $N$  of the sequence to be shortened. The number  $N_e$  of samples that are eliminated by the algorithm is then the smallest multiple of the pitch value  $P$  that is greater than or equal to  $N_{emin}$ . I.e.  $N_e = kP$  where  $k$  is a positive integer and  $N_e \geq N_{emin} > N_e - P$ . The length of the output signal is then  $N_r = N - N_e$ . The input signal to be shortened is written  $s(n)$ , where  $n = 1, \dots, N$  and the output signal is written  $s'(n)$  where  $n = 1, \dots, N_r$ . In order to ensure continuity in the output signal, the first and the last  $N_r$  samples of the signal  $s(n)$  are merged progressively, i.e.  $s'(n) = s(N_e + n) \cdot w(n) + s(n) \cdot (1 - w(n))$  for  $n = 1, \dots, N_r$  where  $w(n)$  is a weighting function such that  $0 \leq w(n) \leq 1$ , for  $n = 1, \dots, N_r$  and  $w(n) \leq w(n+1)$  for  $n = 1, \dots, N_r - 1$ . For example,  $w(n)$  can merely be the linear function  $w(n) = n/N_r$ . For an unvoiced signal where it is not possible to determine the pitch,  $N_e$  can be fixed freely.

Figure 4 showing signal sequences A, B, C, and D illustrates how the process is implemented on a worked example. The first sequence (A) is shown as a solid line and constitutes a piece of the signal  $s(n)$  to be shortened that is  $N = 640$  samples long. The purpose is to shorten this sequence by at least 320 samples without

changing its fundamental frequency, and without introducing any discontinuity or other audible degradation. The pitch of  $s(n)$  varies slowly, its value being equal to 49 at the beginning of the sequence and 45 at the end of the sequence. The pitch detected by a correlation method is  $P = 47$ . Thus,  $s(n)$  will be shortened by  $k = 7$  periods, i.e.  $N_e = kP = 7 \times 47 = 329$  samples.

In this example, linear weighting has been selected. The sequences B and C show two pieces of the signal of length  $N_r = N - N_e = 311$  that have already been weighted and that are subsequently merged together. Merging is performed by adding these two signals together. In sequence C, it can be seen that because of the slight variation in pitch, these two pieces of the signal  $s(n)$  are somewhat phase-shifted. Because of the merging technique used, this does not give rise to any discontinuity in the output signal  $s'(n)$  (continuous line in sequence D). It can also be seen in sequence D that the shortened signal  $s'(n)$  remains properly in phase with the signals that precede it and that follow it (dashed line in Figures 1 and 4).

10019550.032902



## CLAIMS

1/ A method of managing the decoding and playback (14) of a sound signal in an asynchronous transmission system, in which any overabundance of filling of a first buffer memory (11) and/or of a second buffer memory (13) situated at the inlet or at the outlet of a decoding block (12) is detected by comparing the filling level with at least one threshold, the method being characterized in that, depending on the value of the filling level:

- voice activity detection is implemented and frames considered by said detection as being non-active are eliminated; and

- concatenation processing is implemented on two successive frames to compact them into a pseudo-frame of length less than or equal to one frame, the length reduction ratio of the pseudo-frame relative to the length of the two frames being greater than or equal to two.

2/ A method according to claim 1, characterized in that voice activity detection is implemented and frames considered by said detection as being not active are eliminated whenever the filling level lies between a first threshold and a second threshold, and in that concatenation processing is implemented on two successive frames whenever the filling level lies between a second threshold and a third threshold.

3/ A method according to claim 2, characterized in that the first and second thresholds are the same.

4/ A method according to any preceding claim, characterized in that detection is performed at the inlet or the outlet of a decoding block (12) having a first buffer memory (11) at its inlet and/or its outlet to determine whether any frame is missing or erroneous or

CORRECTED SHEET

whether any samples to be played back are absent, and a fake frame is generated to ensure continuity in the audio playback on detecting such a missing or erroneous frame, or on detecting such an absence of samples for playback.

5/ A method according to claim 4, characterized in that when the decoding block (12) implements its decoding processing in cyclical manner relative to the content of the first buffer memory (11), detection of any missing or erroneous frame or of any absence of samples to play back is implemented at the same cyclical frequency, said detection taking place far enough in advance relative to the decoding process to make it possible to generate a fake frame in good time.

6/ A method according to claim 4 or claim 5, characterized in that a fake frame is not generated when a missing or erroneous frame is detected for a frame on which an absence of samples has already been detected.

7/ A method according to any one of claims 4 to 6, characterized in that, for a system of the type which can voluntarily stop sending frames, the type of the previously-generated frame is stored from one frame to the next, and this information is used to determine whether to generate fake frames or to generate frames of silence.

8/ A method according to any preceding claim, characterized in that in processing for concatenating two successive frames, the samples are weighted in such a manner as to give more importance to the first samples of the first frame and to the last samples of the second frame.

9/ A method according to any preceding claim, characterized in that the threshold(s) is/are adaptive.

CORRECTED SHEET

10/ A method according to claim 9, characterized in that a threshold is adapted as a function of the length of time passed with a filling level above a given threshold.

11/ A device for playing back a speech signal, the device comprising a first buffer memory (11) receiving coded frames, means implementing decoding processing (12) on the frames stored in said first buffer memory (11), a second buffer memory (13) receiving decoded frames output by the decoding means, and sound playback means (14) receiving the frames output by the second buffer memory (13), the device being characterized in that it further comprises means for implementing the method according to any preceding claim.

10019550.032902

CORRECTED SHEET

This translation of an amended page covers the amendments made in the original. However, the page breaks match the translation, so that this page is also a replacement page that fits in with the remainder of the translation.

## A B S T R A C T

A METHOD OF MANAGING THE DECODING AND PLAYBACK OF A SOUND  
SIGNAL IN AN ASYNCHRONOUS TRANSMISSION SYSTEM

5

A method of managing the decoding and playback of a  
sound signal in an asynchronous transmission system, in  
which method any overabundance of the filling of said  
buffer memory and/or of a second buffer memory at the  
inlet or the outlet of the decoding block is detected by  
comparing the filling level with at least one threshold,  
the method being characterized in that, depending on the  
value of the filling level:

- voice activity detection is implemented and frames  
considered by said detection as being non-active are  
eliminated; and

- concatenation processing is implemented on pairs  
of successive frames.

20

25

30

Translation of the title and the abstract as they were when originally filed by the  
Applicant. No account has been taken of any changes that may have been made  
subsequently by the PCT Authorities acting ex officio, e.g. under PCT Rules 37.2,  
38.2, and/or 48.3.

A METHOD OF MANAGING THE DECODING AND PLAYBACK OF A SOUND  
SIGNAL IN AN ASYNCHRONOUS TRANSMISSION SYSTEM

The present invention relates to a method of  
managing asynchrony in audio transmission.

5

**GENERAL DESCRIPTION OF THE FIELD OF THE INVENTION**

In general, the invention relates to transmission  
systems using low data rate speech encoders, in which the  
signals do not carry the reference clock of the source  
encoding system (the sampling frequency of the encoder).  
This applies, for example, to Internet protocol (IP) type  
transmissions or indeed to discontinuous transmissions,  
etc.

A general aim of the invention is to resolve the  
problems encountered by such systems in producing a  
continuous stream of decoded speech or sound.

Traditionally, telephone communications and sound  
channel networks have used analog transmission systems  
with frequency division multiplexing (primary groups,  
amplitude and frequency modulation). Under such  
conditions, the speech signal (or music signal; the term  
"speech" is used below throughout this document in  
generic manner) is converted into an electrical signal by  
a microphone and it is this analog signal which is  
filtered and modulated in order to be presented to a  
receiver which amplifies it prior to presenting it to a  
playback system (earphone, loudspeaker, etc.).

Over the last few years, digital transmission and  
switching techniques have been progressively replacing  
analog techniques. In pulse code modulation (PCM)  
systems, the speech signal is sampled and converted into  
a digital signal using an analog-to-digital converter  
(ADC) driven at a fixed sampling frequency derived from a  
master clock delivered by the network and also known to  
the receiver system. This applies to analog and digital  
subscriber connection units in telecommunications  
networks. The digital signal received by the destination

(in the broad sense) is converted back into analog so that it can be heard by means of a digital-to-analog converter (DAC) driven by a clock at the same frequency as that used by the ADC of the source. Under such conditions, the entire system is entirely synchronous as generally applies to present-day switching and transmission systems. These can include data rate reduction systems (for example for a telephone signal, for converting from 64 kilobits per second (kbit/s) to 32 kbit/s or 16 kbit/s or 8 kbit/s). It is the network (or terminal systems as in the case of the integrated services digital network (ISDN) for example) which undertakes the operations of ADC, of encoding and decoding (where encoding and decoding are used in the context of reducing data rate), and of DAC. The clocks are always distributed and the system comprising ADC, speech encoding, transmission and switching, speech decoding, and finally DAC is fully isochronous. There are no losses or repeats of speech samples in the decoder.

The above-described synchronous transmission systems require the presence of a reference clock throughout the network. Transmission systems are now making greater and greater use (initially for data) of asynchronous and packet techniques (IP protocol, asynchronous transmission mode (ATM)). In numerous new situations, the decoder has no reference concerning the sampling frequency used by the encoder and it must be capable, using its own means, of reconstituting a decoding clock which attempts to track the reference of the encoder. The present invention is thus particularly advantageous in frame relay telephone systems, in ATM telephony, or in IP telephony. The technique described can easily be used in other fields of speech or sound transmission in which there exists no effective transmission of the clock reference from the encoder to the decoder.

## DESCRIPTION OF THE STATE OF THE ART

### The general problem

5 The general problem posed by transmission systems to which the invention applies is that of mitigating the fact that the speech or sound decoder has no clock reference associated with the source encoding.

In this respect, two circumstances can be distinguished: those corresponding to "weak" asynchrony and those corresponding to "strong" asynchrony.

10

#### *Weak asynchrony*

As an illustration, we consider the case of a transmission system comprising the following, as shown diagrammatically in Figure 1:

15

- an encoding source 1 comprising an analog-to-digital converter driven by a reference clock at frequency  $F_{ADC}$  equal to 8 kilohertz (kHz) (to provide numerical values in the worked examples below) and a speech encoder (of greater or lesser complexity and reducing the data rate for transmission to a greater or lesser extent);

20

- an asynchronous transmission system (represented by link 2) which conveys the information produced by the encoding source using its own transmission clock and its own protocols (for example the speech encoder could produce a data rate of 8 kbit/s and the transmission system could be constituted by an RS.232 type asynchronous link operating at 9600 bit/s); and

25

- a reception and decoding system 3 receiving the information conveyed over the asynchronous link (whose data rate must necessarily be a little greater than the raw encoding data rate, e.g. 9600 bit/s instead of 8000 bit/s) having the function of producing the signal after decoding (decompression) and applying the signal it produces to a digital-to-analog converter connected to a transducer such as a loudspeaker, a telephone handset, a

30

35

10049550-032593

headset, or a sound card installed in a personal computer (PC).

It will be understood that since the reception and decoding system 3 has no clock reference, it must  
 5 implement a strategy in order to mitigate this asynchrony between the encoder and the decoder.

Whatever the encoding technique used or the type of transmission which does not directly convey a clock, or time markers within the transmitted frame, or indications  
 10 concerning transmission instants, the above-mentioned problem can be reduced (ignoring the speech encoder, the asynchronous transmission system, and the speech decoder) to a system comprising the following, as shown in Figure 2:

- 15       - an analog-to-digital converter 4 for converting speech signals or sound from analog to digital form at a sampling frequency set by a local oscillator;
- a digital-to-analog converter 5 for playing back the sound via a transducer suitable for the field of use  
 20 in question and operating at a sampling frequency given by a local oscillator which, a priori, is at the same frequency but which is never at exactly the same frequency for reasons of acceptable manufacturing cost (highly stable and very accurate frequency sources do  
 25 indeed exist, but they need to be temperature-compensated and they are unacceptably expensive for mass-produced industrial implementation); and
- a digital register 6 into which the analog converter 4 writes at its own sampling frequency ( $F_{ADC}$ ),  
 30 said register being read at the sampling frequency ( $F_{DAC}$ ) of the playback system by the digital-to-analog converter (DAC).

It will be understood that since the two clock frequencies ( $F_{ADC}$  and  $F_{DAC}$ ) are different, it is necessary  
 35 from time to time for the DAC to reread the same information twice over (if  $F_{DAC}$  is greater than  $F_{ADC}$ ) or on



the contrary (where  $F_{DAC}$  is less than  $F_{ADC}$ ) to allow the ADC to overwrite information before the DAC can read it.

The oscillators that are commonly available in the trade are characterized by the accuracy with which they operate (within a certain temperature range).

Oscillators that are accurate to within 50 parts per million (ppm) are quite commonly available and are used to provide numerical values for the worked examples below showing how frequently samples are lost or repeated when the sampling frequency is 8 kHz (the reader can easily determine that at higher sampling frequencies samples are skipped or repeated at a frequency which is prorata the sampling frequency; the higher the sampling frequency the higher the frequency at which samples are skilled or repeated).

Under the least favorable conditions, an ADC is operating at  $8000 \times (1 + 50.e-6)$  in association with a DAC operating at  $8000 \times (1 - 50.e-6)$ . In this particular example, the skip period (period for samples being omitted in the DAC since  $F_{DAC}$  is less than  $F_{ADC}$ ) is easily calculated by counting the number of periods of the DAC (where the period is longer than that of the ADC) that produces a value equal to said period of the DAC when multiplied by the difference between the periods.

Writing the period of the DAC as  $P_{DAC}$  (in this case  $1/8000 \times (1 - 50.e-6)$ ) and  $P_{ADC}$  as the period of the ADC (in this case  $1/8000 \times (1 + 50.e-6)$ ) we obtain  $N \times (P_{DAC} - P_{ADC}) = P_{DAC}$ .  $N$  represents the number of individual operations that stem from the period difference. Writing  $50.e-6 = \epsilon$  and applying the simplifications that are common for small quantities, we obtain  $N = 1/(2\epsilon)$ . In this example, that immediately gives the skip period as being close to 1.25 seconds (s). If the accuracy of the local oscillators is improved (e.g. by going from 50.e-6 to 5.e-6) then the skip period will increase (in this case there will be one skip every 12.5 s).

In a complete transmission system including audio encoders operating on signal frames, this phenomenon of "slip" between two clocks will give rise to an absence of speech frames (no frame to be decoded in the time available for decoding) or to overabundance of frames (i.e. two frames for decoding instead of one in the available time). Taking the example of a speech encoder operating on 30 millisecond (ms) frames at 8 kHz, i.e. 240 samples, in each 30 ms time slot the receiver and more particularly the decoder expects to receive one frame for decoding in order to ensure that playback of the speech signal remains continuous. Unfortunately, if  $F_{ADC}$  is less than  $F_{DAC}$ , then on the above assumptions, there will be an absence of any frame of samples for decoding by the sound playback system once every  $240 \times 1.25 = 300$  s, and in the converse situation there will be two frames instead of one (i.e. a frame to be "eliminated") at the decoder once every 300 s. Under such circumstances, the awkward phenomenon of samples being skipped or repeated becomes very disagreeable since an entire block of the signal is skipped or repeated, and this needs to be managed appropriately.

### *Strong asynchrony*

Certain types of transmission amplify this problem of asynchrony due to the phenomenon of "slip" between clocks as explained above. This is what we refer to herein as "strong" asynchrony.

When transmission is imperfect, giving rise to samples or frames of samples being lost and also when transmission generates jitter on sample arrival times, where such jitter is associated neither with the sending clock nor with the receiving clock, but is associated with other mechanisms in the transmission system having their own clocks, then the receiver system can be confronted with an absence of several frames, or with an overabundance of several frames. This can apply, for

example, with IP type networks which suffer from the phenomenon of packets being lost and from the phenomenon of jitter introduced during packet routing. These phenomena disturb the continuity of the sound playback of the audio signal very strongly. When packets are lost or when jitter delays one or more packets, the playback system finds itself without any sample (or frame of samples) to apply to the DAC for the purpose of ensuring continuity in audio playback. Conversely, when jitter is strong, the playback system can find itself with far too many frames or samples to be sent simultaneously to the DAC. When jitter is strong, sound signal frame transmission can take place in the form of bursts, thus creating phenomena both of gaps and of overabundance amongst sample frames.

It will be observed that using speech encoders operating with a system of the voice activity detector/discontinuous transmission/comfort noise generation (VAD/DTX/CNG) type, a mechanism is also introduced that is similar to the loss of a packet since in the event of silence, the sender will cease to send frames of samples. Ceasing to send samples can be perceived at the receiver as being the same as the loss of a packet or as circumstances in which the ADC clock is faster than the DAC clock, which leads to holes in the signal at the receiver, as shown above.

"Strong" asynchrony thus differs from "weak" asynchrony by involving not only cyclical skips and/or repetitions, but also holes in the signal and/or overabundance of the signal in multiple and non-cyclic manner.

#### **Description of various existing methods**

Two main methods are presently known for mitigating the drawbacks due to the fact that the speech or sound decoder has no clock reference.

10019550 032900

The first consists merely in proceeding as described above in the paragraph describing "weak" asynchrony, i.e. by skipping or repeating samples. The decoding system produces samples at a rate that is more or less equal to that of the encoder and it presents them to the digital-to-analog converter at said rate (means for implementing the above reconstruction system are known to the person skilled in the art). In some cases, for example when "strong" asynchrony applies with transmission being in the form of frames, it is preferable when samples for playing back are missing to send null sample frames to the ADC rather than repeating a preceding frame. Furthermore, in the converse situation, when surplus samples are present, they are not eliminated directly, but a first-in-first-out (FIFO) register of some size can be used to absorb jitter to some extent. If the FIFO register becomes too full, then that triggers partial or complete emptying of the FIFO, thereby giving rise to new skips in sound playback.

20       The second method, which is more complicated and provides better performance, requires a loop to be implemented to recover a hardware clock which is servo-controlled by the filling level of a buffer memory for the signal to be decoded (or to be transmitted as in the ATM adaptation layer number 1 (AAL1) for example). That method of servo-control attempts to use the clock recovery loop to recover the sampling frequency of the source. The filling level of the receive buffer produces a control signal for servo-controlling a digital or analog phase-locked loop (PLL).

30       The first method is extremely simple to implement but suffers from a major defect associated with the quality of the sound reproduced. A skip or elimination once every 1.25 seconds can be very disagreeable to listen to, and this can occur with "weak" asynchrony associated with correction at sample level. Similarly, for a system operating with frames of samples, the

inserted repetitions or blanks, and the discontinuities in the signal due to frames being eliminated amplify loss of quality which becomes highly perceptible and very disturbing for the listener.

5           Furthermore, the use of a FIFO memory runs the risk of establishing a considerable delay in transmission and that also harms the overall quality of a call.

10           The second method is much more complex to implement and requires a clock servo-control mechanism, and thus requires special hardware. However, it provides partial synchronization and therefore avoids problems associated with managing asynchrony. Nevertheless, that method adapts poorly to discontinuous transmission systems, to systems involving last frames, or to systems with high  
15           levels of jitter. Under such circumstances, synchronization information is no longer available. Furthermore, that method cannot be envisaged on terminal platforms where clock servo-control is not possible, as is the case in particular with PC type terminals, for  
20           example, where the system used for playing back sound is a sound card.

#### **SUMMARY OF THE INVENTION**

25           A general object of the invention is to propose a solution to the problems associated with continuity in the playback of a speech signal in the presence of asynchronous transmission, and to do so by taking action at receiver level, i.e. at the end of the transmission system.

30           To this end, the invention provides a method of managing the decoding and playback of a sound signal in an asynchronous transmission system, in which method any overabundance of the filling of said buffer memory and/or of a second buffer memory at the inlet or the outlet of  
35           the decoding block is detected by comparing the filling level with at least one threshold, the method being

characterized in that, depending on the value of the filling level:

- voice activity detection is implemented and frames considered by said detection as being non-active are eliminated; and

- concatenation processing is implemented on pairs of successive frames.

Such a method is simple to implement and provides a guarantee of quality by avoiding excessive increase in transmission delay and by managing holes in the speech signal effectively. Furthermore, it does not imply any specific hardware servo-control circuit, and can therefore be quickly adapted to different asynchronous networks, terminals, and platforms.

The method is advantageously associated with the various characteristics below taken singly or in any technically feasible combination:

- voice activity detection is implemented and frames considered by said detection as being not active are eliminated whenever the filling level lies between a first threshold and a second threshold, and in that concatenation processing is implemented on two successive frames whenever the filling level lies between a second threshold and a third threshold;

- the first and second thresholds are the same;

- detection is performed at the inlet or the outlet of a decoding block having a first buffer memory at its inlet and/or its outlet to determine whether any frame is missing or erroneous or whether any samples to be played back are absent, and a fake frame is generated to ensure continuity in the audio playback on detecting such a missing or erroneous frame, or on detecting such an absence of samples for playback;

- when the decoding block implements its decoding processing in cyclical manner relative to the content of the first buffer memory, detection of any missing or erroneous frame or of any absence of samples to play back

is implemented at the same cyclical frequency, said detection taking place far enough in advance relative to the decoding process to make it possible to generate a fake frame in good time;

5           - a fake frame is not generated when a missing or erroneous frame is detected for a frame on which an absence of samples has already been detected;

          - for a system of the type which can voluntarily stop sending frames, the type of the previously-generated  
10 frame is stored from one frame to the next, and this information is used to determine whether to generate fake frames or to generate frames of silence;

          - in processing for concatenating two successive frames, the samples are weighted in such a manner as to  
15 give more importance to the first samples of the first frame and to the last samples of the second frame;

          - the threshold(s) is/are adaptive; and

          - a threshold is adapted as a function of the length of time passed with a filling level above a given  
20 threshold.

The invention also provides a device for playing back a speech signal, the device comprising a first buffer memory receiving coded frames, means implementing decoding processing on the frames stored in said first  
25 buffer memory, a second buffer memory receiving decoded frames output by the decoding means, and sound playback means receiving the frames output by the second buffer memory, the device being characterized in that it further comprises means for implementing the above-specified  
30 method.

As will be understood on reading the following description, these means are essentially computer means.

#### DESCRIPTION OF THE FIGURES

35           Other characteristics and advantages of the invention appear further from the following description which is purely illustrative and non-limiting and which

should be read with reference to the accompanying figures, in which:

- Figure 1 is a block diagram of an asynchronous transmission system;
- 5       - Figure 2 is a diagram showing a model of such a transmission system;
- Figure 3 is a diagram of a receiver device; and
- Figure 4 shows the signals obtained by implementing concatenation processing as proposed by the invention.

#### DETAILED DESCRIPTION OF ONE OR MORE EMBODIMENTS

The method proposed by the invention for managing asynchrony in transmission implements two kinds of processing corresponding to handling the two phenomena described above, namely the lack of samples and surplus samples.

##### 1. Description of the sound playback system in a conventional transmission application

As shown in Figure 3, the playback system for a speech signal comprises three elements:

- A block 10 waiting to receive samples or frames of code coming from the network. The block 10 contains a FIFO type memory 11 or circular type buffer memory (referred to as "FIFO 1" in the description below) enabling frames to be stored on a temporary basis prior to being decoded.
- A decoding block 12 which takes the frames coming from the block 10, decodes them, and stores them in turn in a FIFO memory 13 (referred to below as "FIFO 2").
- A playback block 14 which takes the decoded sample frames and applies them to any kind of sound playback system 15.

Depending on the terminals and the way the system is organized, the clock frequency used for sound playback (i.e. the digital-to-analog converter frequency  $F_{DAC}$ ) is



not necessarily directly associated with all of the blocks. Since the block 14 is directly associated with the playback system, it is directly associated with the frequency  $F_{DAC}$ . However the other blocks can be  
 5 associated instead with the rate at which frames arrive from the network rather than with the frequency  $F_{DAC}$ . Taking the example of a terminal provided with a multitasking system, and in which each block is performed by a specific task, the tasks 10 and 12 can thus be  
 10 associated with frame reception. The task 10 waits for a frame from a network, which frame is then decoded by the task 12 and placed in the memory FIFO 2.

Meanwhile the task 14 clocked at  $F_{DAC}$  takes samples from the memory FIFO 2 and delivers them continuously to  
 15 the sound playback system.

It can thus be seen that regardless of whether the asynchrony is "strong" or "weak", it is the way in which the memory FIFO 2 is managed that requires particular attention. Similarly, if the task 12 were strongly  
 20 associated with the task 14, then particular attention would be required by the memory FIFO 1.

The mechanism constituting an implementation of the invention is described below in application to managing the memory FIFO 2, but the description includes  
 25 explanations about how to transpose it with certain adaptations to managing the memory FIFO 1.

## 2. Absence of samples

In order to continue playing back sound in the  
 30 absence of samples, both potential causes of samples for playback being absent are treated. The first cause corresponds to information contained in lost packets, while the second cause corresponds to the absence of any samples to play back (e.g. FIFO 2 empty) even though it  
 35 is still necessary to keep on sending samples to the sound playback system.

## 2.1 Loss of frames or erroneous frames

The processing applied to lost frames or to erroneous frames requires a transmission system to be available that gives access to information about frames being lost and about erroneous frames being received. This is often the case in transmission systems.

For example, in IP networks, it is possible to use the marking of packets coming from the real time transfer (RTP) layer, which marking gives the exact number of samples lost between two packets of audio code being received. This information about loss of frames, or in the case of IP about loss of packets (each containing one or more speech frames) generally becomes available only once the packet following the lost packet(s) is itself received.

It is not necessarily advantageous to take action, while one or more valid frames can be decoded. With new generation speech encoders (CELP encoders, transform encoders, ...), in order to ensure that the quality of sound playback is maintained, it is often necessary to ensure a degree of synchronism between the encoder and the decoder. The loss of this encoder/decoder synchronism can be compensated by using frame loss correction algorithms associated with the speech encoder used. By way of example, these algorithms are provided in the standards for certain speech encoders (e.g. International Telecommunications Union (ITU) standard G.723.1). When using simpler encoders, such a mechanism is not always necessary.

When a large number of frames have been lost, the number of "fake" sample frames that need to be generated in order to pack out the memory FIFO 2 can be limited. The purpose of processing fake frame generation is to fill holes in such a manner as to ensure signal continuity while also smoothing the internal variables of the decoder so as to avoid excessive divergence on decoding the first valid frame following the invalid or

lost frames, thereby avoiding any audible discontinuity. After a few frames have been generated, it can be assumed that the variables have been smoothed, and thus that the generation of such fake frames can be limited to a small number of frames (e.g. four to six) whenever a large  
 5 number of frames has been lost.

It will thus be understood that processing is servo-controlled in this way relative to information about lost frames.

10 Similar processing is implemented on the basis of information about invalid frames. This information is forwarded to the decoder by the network portion of the receiver and it arrives soon enough to enable a frame correction algorithm to be implemented which, by taking  
 15 account of such a non-valid frame, makes it possible to ensure continuity in the signal, and thus to avoid having another cause for samples being absent in the memory FIFO 2.

To sum up, this first process corresponds to  
 20 managing information of the type "n frames lost" or "invalid frame received" coming from the network layer of the receiver. This management is characterized by implementing an algorithm for correcting frame losses (also referred to in this document as an algorithm for  
 25 generating "fake" frames). This first process therefore acts at decoding task level and feeds the memory FIFO 2.

## 2.2 Absence of samples to be played back

This second process is associated with the clock coming from the task 14, i.e. with the clock at the  
 30 frequency  $F_{DAC}$ . As mentioned above, the memory FIFO 2 (or FIFO 1 if the task 12 is included in the task 14) can become empty of samples even though it is still necessary to supply samples to the sound playback system. It is  
 35 then necessary to supply the playback system with samples and if possible to avoid playing back zeros (since this degrades the sound signal very greatly).

This second process can be thought of as a feedback loop on frame decoding. This loop causes the algorithm for correcting frame losses to be called and as a result it needs to be activated soon enough to enable the  
 5 algorithm to be executed and to enable samples to be sent to the sound playback system. Depending on the platform, this feedback can be called in different ways.

This loop can be implemented in two ways which are described below.

10 For a single-task receiver (e.g. a digital signal processor (DSP) without any real time operating system (RTOS)), the audio decoder portion is tied completely to the DAC clock ( $F_{DAC}$ ) and is therefore permanently waiting for a frame to be decoded in cyclical manner. For  
 15 example, with a speech encoder using 30 ms frames, waiting loops are built up in periods that are multiples of 30 ms.

Thus, for a 30 ms loop, the decoder will, every 30 ms, be expecting a frame for decoding to be placed in the memory FIFO 1 (which can correspond merely to a frame passing from the network layer to the task 12). On  
 20 arrival of the frame, it is decoded and placed in the memory FIFO 2 for sending to the DAC. The feedback processing is implemented whenever it is observed that there is no frame for decoding in the memory FIFO 1 at  
 25 the time given by:

$$T = T_o + 30 \text{ ms} - T_c$$

where:

30  $T_o$  = the start time of the 30 ms wait loop; and  
 $T_c$  = the time required for executing the algorithm for generating fake frames with a safety margin corresponding to interrupts and/or other auxiliary processing that might take place before the end of the loop.

35 Processing is thus implemented with a latency time deadline of  $T_b$  (loop time) -  $T_c$  (computation time + margin).

5

10

15

25

30

35

system by implementing a check on the content of the  
buffer memory FIFO 2 (or FIFO 1 depending on how the  
sound playback system is managed) and activates the  
algorithm for generating "fake" frames in order to  
5 generate the missing samples.

It will be understood that the second process  
responds firstly to the problem of "slip" between clocks,  
and more precisely to the circumstance in which the  
received clock ( $F_{DAC}$ ) is faster than the send clock ( $F_{ADC}$ ).  
10 It also applies to the phenomenon of frames being lost  
since this can lead to there being an absence of samples  
to send to the DAC even before frame loss has been  
detected, since such detection occurs only on receiving  
the frame following the loss.

15 In order to combine the actions of the first and  
second processes, the first process is prevented from  
generating "fake" frames on detection of frame loss  
whenever the corresponding frames have just been  
generated by the second process.

20 For this purpose, use is made of flags and also of  
counters determining the number of samples that have been  
generated by the second process.

### 2.3 Specific actions for speech encoders using 25 VAD/DTX/CNG services

Encoders using a VAD/DTX/CNG system can voluntarily  
stop sending frames; under such circumstances, the  
absence of samples must not be considered exactly as a  
loss of frames, but rather as a period of silence. The  
30 only way of determining whether the frame to be generated  
must be silence or should correspond to a lost frame is  
to know the type of the previously-generated frame (i.e.  
signal frame or frame corresponding to a lost frame, or a  
noise update frame (SID), or a frame of silence (NOT)).  
35 For this purpose, the type of the generated frame is  
stored, and while frames are being generated to  
compensate for an absent frame or a lost frame, it is

10019550 032900

decided whether fake frames should be generated using the algorithm for correcting frame losses (as applies when the preceding frame was of the FSF type), or whether frames of silence should be generated by activating the decoder appropriately (as applies when the preceding frame was of the SID or the NOT type).

### 3. Overabundance of samples to be played back

In order to deal with an overabundance of samples to be played back, processing is implemented to empty out frames, eliminating certain frames in full or in part prior to their possibly being taken into account by the sound playback system.

This processing enables frames to be stored in buffer memories until certain thresholds trigger actions for limiting the amount of frames in memory and for limiting any corresponding increase in delay across the communications system. This limited storage makes it possible to accommodate jitter phenomena on receiving frames in bursts and also slip between clocks, while nevertheless limiting transmission delay.

#### 3.1 Emptying out processing

Any accumulation of frames is initially detectable in the memory FIFO 1, and is subsequently transferred to the memory FIFO 2.

The proposed method manages information concerning the filling level of a reference buffer memory, i.e. FIFO 1 or FIFO 2 depending on how the tasks 10, 12, and 14 are organized in the receiver (as explained above). If the tasks 12 and 14 are associated with each other, then the filling level information used by the method relates to the memory FIFO 1 which acts as a buffer between the network and the sound playback system. Similarly, if the tasks 10 and 12 are associated, then it is the memory FIFO 2 which acts as a buffer and it is

therefore its filling level which is taken into consideration by the management process.

The process is explained below for the second context. The first is merely an immediate transposition thereof.

5 In order to maintain synchronization as closely as possible between the encoder and the decoder, and thus maintain optimum sound playback, all of the frames coming from the network are decoded. The process then decides  
10 on what action to take on the decoded frame as a function of information concerning filling level. This action is described in greater detail below. To activate the processing, filling level thresholds are used. These thresholds define filling alarm levels for the FIFO  
15 memory. In order to take action that is as inaudible as possible (i.e. in order to limit quality degradation) two levels of action are selected. A first level (alarm level 1) corresponds to the filling level being excessive but not critical (far from the maximum acceptable filling  
20 level), while a second level (alarm level 2) corresponds to it being mandatory to take action on each frame (this level is quite close to the maximum acceptable level). A third or "emergency" level (alarm level 3) is also defined in order to avoid memory overflows or other  
25 problems. This level corresponds to filling being very close to the maximum acceptable level. Alarm level 3 should never be reached if the actions taken at the two preceding threshold levels are properly performed and if the thresholds are properly defined.

30 Each time decoding is performed, the information concerning filling level is compared with the thresholds in order to determine the state of the FIFO (in an alarm state or not), and, where appropriate, the level of the alarm.

35 If the state obtained is not an alarm state, then no action is undertaken and the decoded frame is stored in FIFO 2.



10019550"032900

In the first alarm state, it is considered that at least 50% of the signal coming from a conversation is not useful and therefore at this alarm level, all frames presenting very little information are eliminated. For this purpose, it is possible to implement simple VAD which monitors all frames of samples after they have been decoded to decide whether or not they should be written into FIFO 2. The process can also make decisions based on information taken directly from the code frame concerning the importance or otherwise of the information contained in the frame. In this alarm state, any frame that is considered as containing nothing but noise is simply not stored in FIFO 2 for future sound playback.

In the second alarm state (critical level), it is necessary to take action on each frame to curb growth in the filling level of the memory FIFO 2 very aggressively. At this level, the preceding processing (i.e. the processing implemented for alarm level 1) remains active. However it is now also necessary to shorten pairs of consecutive frames down to the length of one frame or shorter. A decision is therefore taken on the basis of two non-"silent" sample frames (given that any frame that is "silent" is merely not written to FIFO 2 as a result of alarm state 1 being already included in alarm state 2). Action on two consecutive frames is therefore undertaken only when a frame is detected as being non-"silent". The frame is initially stored, and then if the next frame is "silent", then it is only the first frame that is written into FIFO 2.

When both frames contain important information, it becomes necessary to replace them by a single frame while minimizing loss of information and degradation of quality. It is the replacement frame that is stored in FIFO 2. Any effective solution capable of performing this task can be used and activated under such conditions (i.e. second alarm state and two non-"silent" frames).

Two examples of algorithms for performing this task are described below.

In a first algorithmic solution, the two contiguous frames are replaced by a single frame in which each  
 5 coefficient  $x_j$  (where  $j$  lies in the range 0 to  $N-1$  and where  $N$  is the number of samples per frame) is given the value  $(x_i + x_{i+1})/2$  (where  $i$  lies in the range 0 to  $2N-1$ , with the coefficients  $x_i$  coming from both original frames). This solution amounts to performing a kind of  
 10 smoothed undersampling. The frequency of the played-back signal is thus doubled for this frame. Nevertheless, the inventors have found that providing alarm state 2 does not occur very frequently, this solution suffices to maintain the quality of sound playback.

15 In a second solution, signal amplitude is detected to enable the two frames to be compacted into a pseudo-frame of length shorter than or equal to that of one frame. The number of samples contained in the pseudo-frame is determined by the fundamental frequency or  
 20 "pitch" information, but in all events it is shorter than or equal to the length of a normal frame, while nevertheless being close to said frame length. The algorithm used ensures continuity of the played-back signal without any audible hole, and without any  
 25 frequency doubling, while nevertheless dividing the amount of storage required for the signal by a factor that is greater than or equal to 2. This is explained in greater detail in paragraph 3.4 below. Furthermore, this also minimizes the loss of sound information since less  
 30 than 50% of the information is in fact eliminated.

It should be observed that when the receiver implements its processing on the basis of analyzing  
 FIFO 1, with the decoder being directly associated with the sound playback system, it is necessary to generate a  
 35 number of samples that is sufficient, i.e. in the present case a number that ensures at least one frame of samples is made available in FIFO 2. The frame concatenation

algorithm is then calibrated to ensure that it always generates a minimum number of samples, but at least one frame. Another solution would consist in activating the algorithm several times over instead of only once when  
 5 that is necessary to obtain a sufficient number of samples.

In the third alarm level, which is normally never reached, no frames are stored in FIFO 2. In a variant, the system can also decide to clear out a fraction of the  
 10 buffer memory suddenly (this can apply where it is management of FIFO 1 that is activated).

It should also be observed that depending on the network and on the types of problems it generates in terms of asynchrony, it is possible to decide whether or  
 15 not to activate particular alarm levels. For example, when asynchrony is "weak" then alarm levels 1 and 2 can be combined, and the simple solution of replacing two frames by a single frame can then constitute the only active process.

20

### 3.2 Alarm thresholds

There follows a more detailed description of the alarm thresholds and how they are managed.

As explained above, the reference memory is said to  
 25 be in alarm state 1 when its filling level is above threshold 1; this state remains active until its filling level drops below threshold 0. State 1 therefore operates with hysteresis.

The memory is said to be in alarm state 2 if the  
 30 filling level exceeds threshold 2 and to be in alarm state 3 if the filling level exceeds threshold 3. It is possible to envisage managing these alarm states with hysteresis as well.

Thresholds 0, 1, and 2 are adaptive. Threshold 3 is  
 35 directly associated with the maximum acceptable size and it is fixed. These thresholds need to be adaptable in order to accommodate different call contexts and real

10049550-032900  
 206220-05567001

time fluctuations during a call. It is appropriate to allow a greater amount of delay when the call is suffering a large amount of jitter (where delaying playback remains the best way of ensuring acceptable quality in the presence of jitter). In a high-jitter context, it is therefore appropriate for the thresholds 0, 1, and 2 to be at quite high levels.

To facilitate processing, the positions of the thresholds can correspond to integer numbers of frame sizes as exchanged between the various tasks of the receiver. This frame size is written  $T_t$ .

By way of example, the initial values of these thresholds can be as follows:

Threshold 0:  $5 \times T_t$   
 Threshold 1:  $8 \times T_t$   
 Threshold 2:  $12 \times T_t$   
 Threshold 3:  $24 \times T_t$  (fixed value)

The thresholds 0, 1, and 2 can be adapted together in steps of size  $T_t$ . Extreme acceptable values can lie in the range -1 to +8, for example.

Thus, threshold 1 can take on values  $7 \times$ ,  $8 \times$ ,  $9 \times$ ,  $10 \times$ , ...,  $16 \times T_t$ . Threshold adaptation proper is performed on the basis of an adaptation criterion which is the length of time spent in the alarm state. For this purpose, an alarm state percentage is evaluated about once every  $N$  seconds (e.g.  $N = 10$ ). When this percentage exceeds a given threshold (5%), the alarm state thresholds are increased; otherwise, when these percentages are below a given minimum threshold (0.5%) the alarm threshold are decreased. To ensure that the system does not oscillate excessively due to its thresholds being adapted too frequently, hysteresis is applied to adaptation decision making. The thresholds are actually increased by one step, only in the presence of two increase options that are consecutive, and they are decreased by one step, only in the presence of three decrease options that are consecutive. As a result, the length of time between two

threshold increments is at least  $2N$  seconds and the length of time between two threshold decrements is at least  $3N$  seconds. The procedure for increasing thresholds can be accelerated if a large percentage of frames are in an alarm state. One accelerating procedure consists in increasing the thresholds directly, for example whenever the alarm percentage exceeds 50%.

Naturally, the threshold values given for the alarm thresholds are provided purely by way of indication.

### 3.3 Interaction with the first process

The first process is the process which causes "fake" frames to be generated when frames are lost or erroneous. When the system is in an alarm state (overabundance of frames), there is no need to generate "fake" frames which would merely aggravate the phenomenon of overabundance. Nevertheless, in order to maintain high quality sound playback it is important to maintain encoder/decoder synchronization by informing the decoder whenever a frame has been lost (e.g. by launching the generation of one or two fake frames, but no more). The third process will act in the alarm state on the first process so as to curb very strongly the generation of "fake" frames.

### 3.4 Frame concatenation

The purpose of the concatenation process is to shorten the duration of a digital audio signal containing speech or music while introducing as little audible degradation as possible. Since the sampling frequency is given and fixed, it is the number of samples sent to the sound playback apparatus that is decreased. One obvious solution for shortening a sequence of  $N$  samples is to remove  $M$  regularly spaced apart samples from the sequence in question. This causes the fundamental frequency to increase and that can be unpleasant for the listener, particularly when the ratio  $M/N$  is large. Furthermore, there is a danger of no longer complying with the

sampling theorem. The process described below makes it possible to shorten an audio sequence without modifying its fundamental frequency and without giving rise to audible degradation due to signal discontinuity. This process is based on detecting the value of the pitch period. The number of samples eliminated by this algorithm cannot be selected freely, since it is a multiple of the pitch value  $P$ . Nevertheless, it is possible to define a minimum number of samples to be eliminated  $N_{emin}$  which must satisfy the relationship  $N_{emin} \leq N/2$ . Since the purpose is to eliminate at least 50% of the samples in the context of the device for managing asynchrony in an audio transmission, it is advantageous to set  $N_{emin} = N/2$ . It is also assumed that the maximum value of the pitch  $P$  is less than the length  $N$  of the sequence to be shortened. The number  $N_e$  of samples that are eliminated by the algorithm is then the smallest multiple of the pitch value  $P$  that is greater than or equal to  $N_{emin}$ . I.e.  $N_e = kP$  where  $k$  is a positive integer and  $N_e \geq N_{emin} > N_e - P$ . The length of the output signal is then  $N_r = N - N_e$ . The input signal to be shortened is written  $s(n)$ , where  $n = 1, \dots, N$  and the output signal is written  $s'(n)$  where  $n = 1, \dots, N_r$ . In order to ensure continuity in the output signal, the first and the last  $N_r$  samples of the signal  $s(n)$  are merged progressively, i.e.  $s'(n) = s(N_e=n).w(n) + s(n).(1-w(n))$  for  $n = 1, \dots, N_r$  where  $w(n)$  is a weighting function such that  $0 \leq w(n) \leq 1$ , for  $n = 1, \dots, N_r$  and  $w(n) \leq w(n+1)$  for  $n = 1, \dots, N_r-1$ . For example,  $w(n)$  can merely be the linear function  $w(n) = n/N_r$ . For an unvoiced signal where it is not possible to determine the pitch,  $N_e$  can be fixed freely.

Figure 4 showing signal sequences A, B, C, and D illustrates how the process is implemented on a worked example. The first sequence (A) is shown as a solid line and constitutes a piece of the signal  $s(n)$  to be shortened that is  $N = 640$  samples long. The purpose is to shorten this sequence by at least 320 samples without

changing its fundamental frequency, and without introducing any discontinuity or other audible degradation. The pitch of  $s(n)$  varies slowly, its value being equal to 49 at the beginning of the sequence and 45 at the end of the sequence. The pitch detected by a correlation method is  $P = 47$ . Thus,  $s(n)$  will be shortened by  $k = 7$  periods, i.e.  $N_e = kP = 7 \times 47 = 329$  samples.

In this example, linear weighting has been selected. The sequences B and C show two pieces of the signal of length  $N_r = N - N_e = 311$  that have already been weighted and that are subsequently merged together. Merging is performed by adding these two signals together. In sequence C, it can be seen that because of the slight variation in pitch, these two pieces of the signal  $s(n)$  are somewhat phase-shifted. Because of the merging technique used, this does not give rise to any discontinuity in the output signal  $s'(n)$  (continuous line in sequence D). It can also be seen in sequence D that the shortened signal  $s'(n)$  remains properly in phase with the signals that precede it and that follow it (dashed line in Figures 1 and 4).

10019550 032903

## CLAIMS

- 1/ A method of managing the decoding and playback of a sound signal in an asynchronous transmission system, in which method any overabundance of the filling of said  
5 buffer memory and/or of a second buffer memory at the inlet or the outlet of the decoding block is detected by comparing the filling level with at least one threshold, the method being characterized in that, depending on the value of the filling level:
- 10       - voice activity detection is implemented and frames considered by said detection as being non-active are eliminated; and
- concatenation processing is implemented on pairs of successive frames.
- 15       2/ A method according to claim 1, characterized in that voice activity detection is implemented and frames considered by said detection as being not active are eliminated whenever the filling level lies between a  
20 first threshold and a second threshold, and in that concatenation processing is implemented on two successive frames whenever the filling level lies between a second threshold and a third threshold.
- 25       3/ A method according to claim 2, characterized in that the first and second thresholds are the same.
- 4/ A method according to any preceding claim, characterized in that detection is performed at the inlet  
30 or the outlet of a decoding block having a first buffer memory at its inlet and/or its outlet to determine whether any frame is missing or erroneous or whether any samples to be played back are absent, and a fake frame is generated to ensure continuity in the audio playback on  
35 detecting such a missing or erroneous frame, or on detecting such an absence of samples for playback.

10019550 032900



5/ A method according to claim 4, characterized in that when the decoding block implements its decoding processing in cyclical manner relative to the content of the first buffer memory, detection of any missing or erroneous frame or of any absence of samples to play back is implemented at the same cyclical frequency, said detection taking place far enough in advance relative to the decoding process to make it possible to generate a fake frame in good time.

6/ A method according to claim 4 or claim 5, characterized in that a fake frame is not generated when a missing or erroneous frame is detected for a frame on which an absence of samples has already been detected.

7/ A method according to any one of claims 4 to 6, characterized in that, for a system of the type which can voluntarily stop sending frames, the type of the previously-generated frame is stored from one frame to the next, and this information is used to determine whether to generate fake frames or to generate frames of silence.

8/ A method according to any preceding claim, characterized in that in processing for concatenating two successive frames, the samples are weighted in such a manner as to give more importance to the first samples of the first frame and to the last samples of the second frame.

9/ A method according to any preceding claim, characterized in that the threshold(s) is/are adaptive.

10/ A method according to claim 9, characterized in that a threshold is adapted as a function of the length of time passed with a filling level above a given threshold.

- 11/ A device for playing back a speech signal, the device comprising a first buffer memory receiving coded frames, means implementing decoding processing on the frames stored in said first buffer memory, a second buffer  
5 memory receiving decoded frames output by the decoding means, and sound playback means receiving the frames output by the second buffer memory, the device being characterized in that it further comprises means for implementing the method according to any preceding claim.

10019550.000000

## A B S T R A C T

A METHOD OF MANAGING THE DECODING AND PLAYBACK OF A SOUND  
SIGNAL IN AN ASYNCHRONOUS TRANSMISSION SYSTEM

5

A method of managing the decoding and playback of a sound signal in an asynchronous transmission system, in which method any overabundance of the filling of said buffer memory and/or of a second buffer memory at the inlet or the outlet of the decoding block is detected by comparing the filling level with at least one threshold, the method being characterized in that, depending on the value of the filling level:

- voice activity detection is implemented and frames considered by said detection as being non-active are eliminated; and

- concatenation processing is implemented on pairs of successive frames.

20

25

30

Translation of the title and the abstract as they were when originally filed by the Applicant. No account has been taken of any changes that may have been made subsequently by the PCT Authorities acting ex officio, e.g. under PCT Rules 37.2, 38.2, and/or 48.3.

35

"006220" 05561001

1/2

FIG. 1

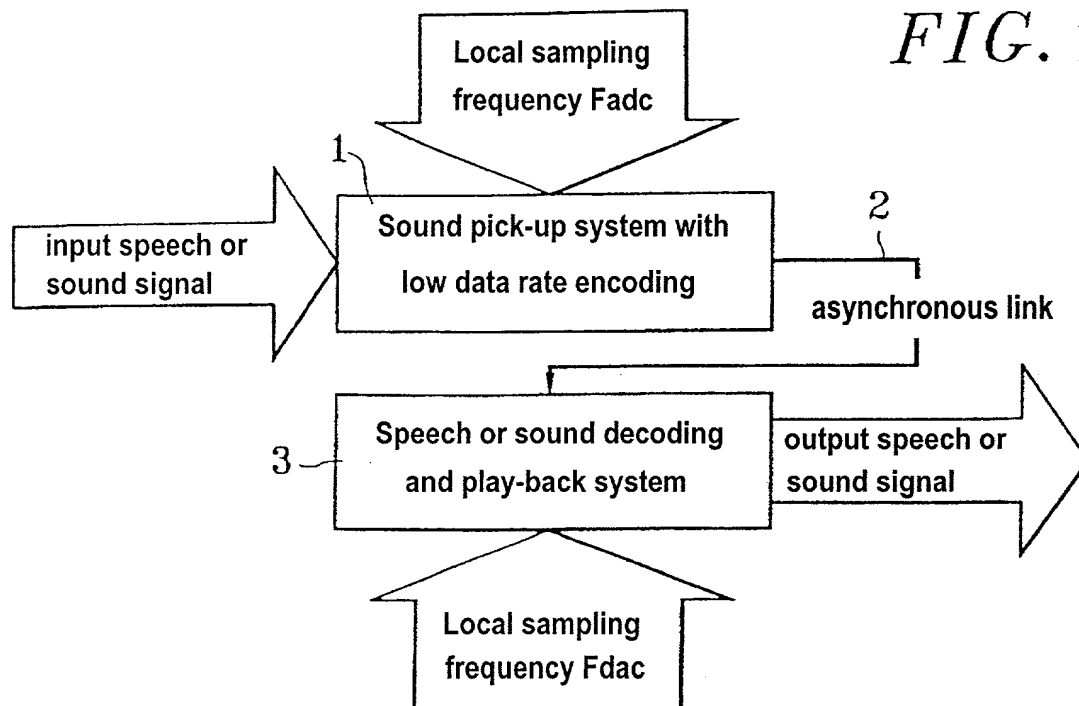
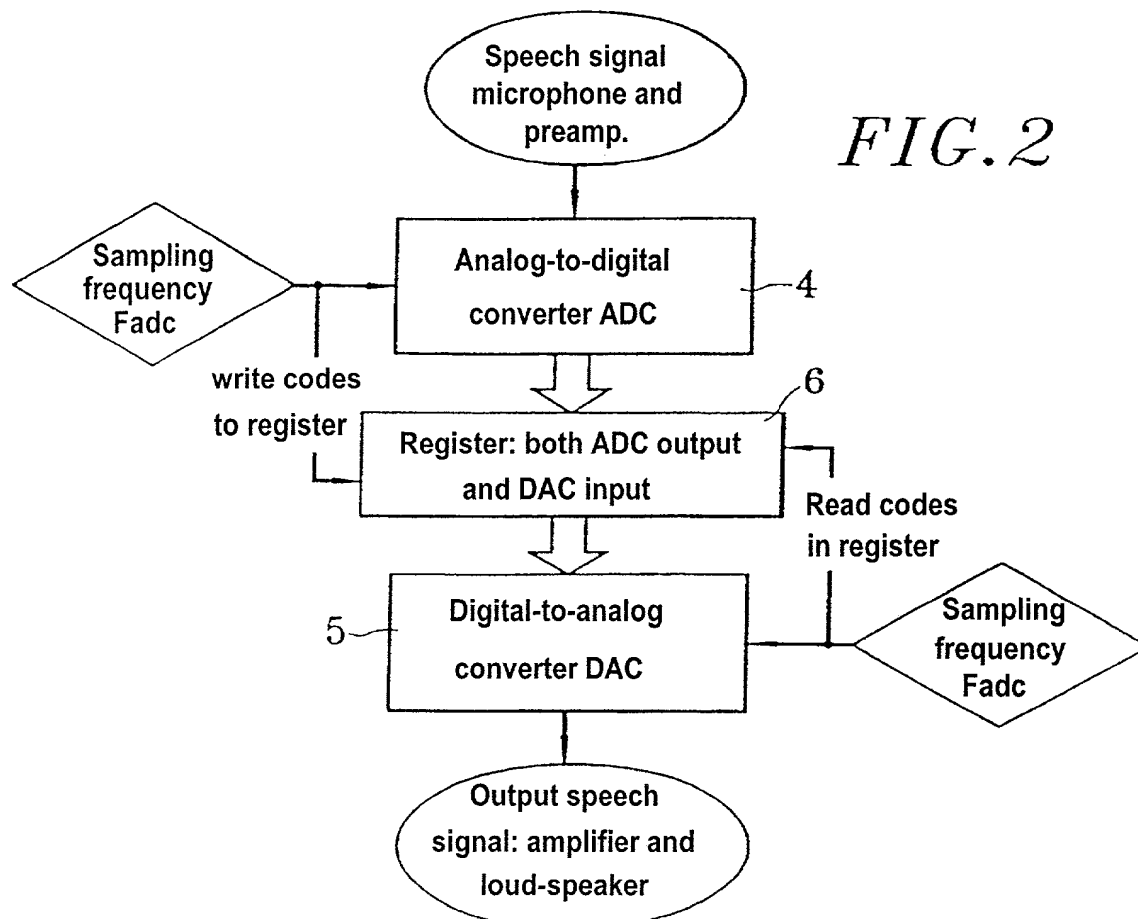


FIG. 2



2/2

FIG. 3

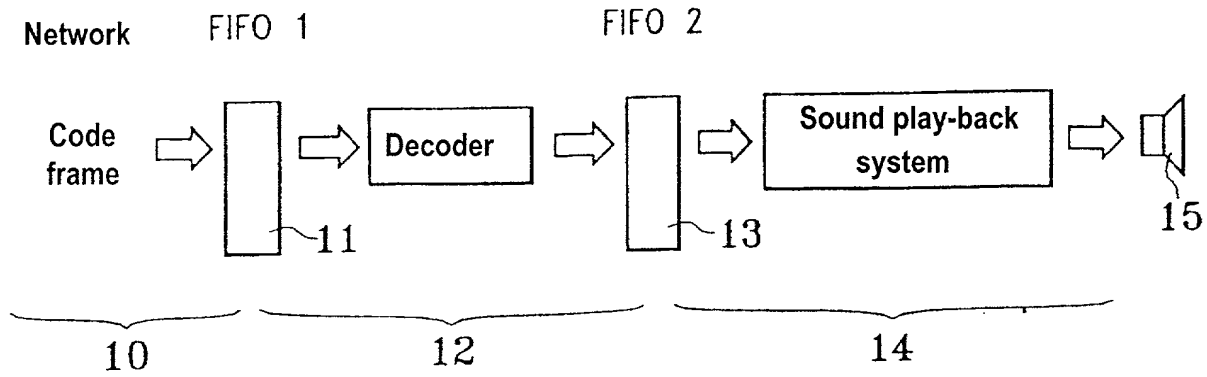
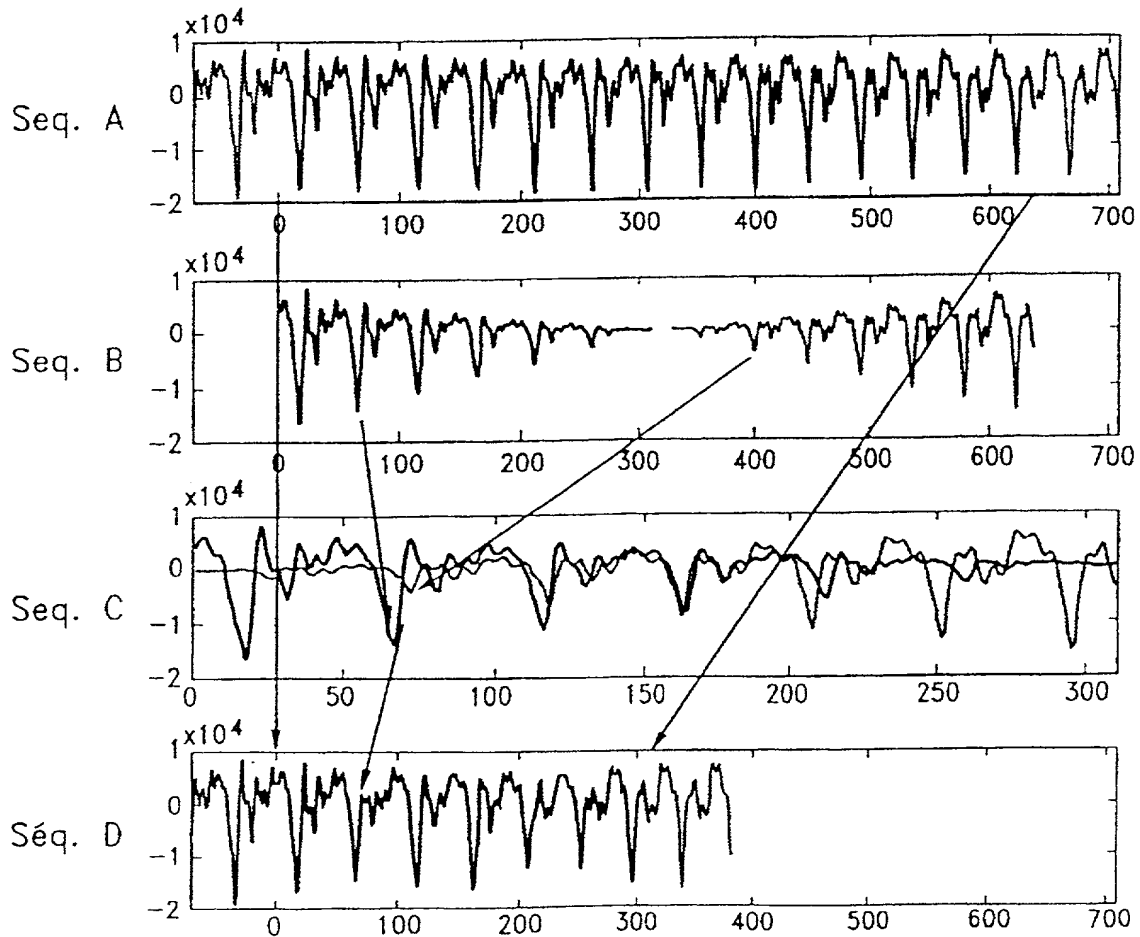


FIG. 4





ref.: 1565.P387

#3

## DECLARATION AND POWER OF ATTORNEY FOR PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below, next to my name,

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

**METHOD FOR DECODING AND RETRIEVING A SOUND SIGNAL IN AN ASYNCHRONOUS TRANSMISSION SYSTEM**

the specification of which

is attached hereto  
was filed on June 22, 2000 as  
Application Serial No. PCT/FR00/01734  
And was amended on  
(if applicable)

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above. I do not know and do not believe that the same was ever known or used in the United States of America before my invention thereof, or patented or described in any printed publication in any country before my invention thereof or more than one year prior to this application, that the same was not in public use or on sale in the United States of America more than one year prior to this application, and that the invention has not been patented or made the subject of an inventor's certificate issued before the date of this application in any country foreign to the United States of America on an application filed by me or my legal representatives or assigns more than twelve months prior to this application.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, Section 1.56(a).

I hereby claim foreign priority benefits under Title 35, United States Code, Section 199, of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor(s) certificate having a filing date before that of the application on which priority is claimed:

Prior Foreign Application(s)			Priority Claimed	
99/08081	FRANCE	24 JUNE 1999	X	
(Number)	(Country)	(Day/Month/Year Filed)	Yes	No
(Number)	(Country)	(Day/Month/Year Filed)	Yes	No
(Number)	(Country)	(Day/Month/Year Filed)	Yes	No

I hereby claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, Section 1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application.

PCT/FR00/01734  
(Application Serial No.)

22 JUNE 2000  
(Filing Date)

Pending  
(Status - patented, pending, abandoned)

(Application Serial No.)

(Filing Date)

(Status - patented, pending, abandoned)

(Application Serial No.)

(Filing Date)

(Status - patented, pending, abandoned)

(42) I hereby appoint BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN, a firm including :

Keith G. Askoff, Reg. No. 33,828; Aloysius T.C. AuYeung, Reg. No. 35,432; Bradley J. Bereznek, Reg. No. 33,474; Michael A. Bernadicou, Reg. No. 35,934; Roger W. Blakely, Jr.; Reg. No. 25,831; Timothy R. Croll, Reg. No. 36,771; Daniel M. De Vos, Reg. No. 37,813; Scott A. Griffin, Reg. No. 38,167; Stephen D. Gross, Reg. No. 31,020; David R. Halvorson, Reg. No. 33,395; Michael D. Hartogs, Reg. No. 36,547; Brian D. Hickman, Reg. No. 35,894; George W. Hoover II, Reg. No. 32,992; Paul H. Hostmann, Reg. No. 36,167; Eric S. Hyman, Reg. No. 30,139; Dag H. Johansen, Reg. No. 36,172; Stephen L. King, Reg. No. 19,180; Joseph T. Lin, Reg. No. 38,225; Michael J. Mallie, Reg. No. 36,591; James D. McFarland, Reg. No. 32,544; Anthony C. Murabito, Reg. No. 35,295; Kimberley G. Nobles, Reg. No. 38,255; Ronald W. Reagin, Reg. No. 20,340; Kent R. Richardson, Reg. No. P-39,443; James H. Salter, Reg. No. 35,668; William W. Schaal, Reg. No. P-39,018; James C. Sheller, Reg. No. 31,195; Edward W. Scott IV, Reg. No. 36,000; Maria E. Sobrino, Reg. No. 31,639; Stanley W. Sokoloff, Reg. No. 25,128; Allan T. Sponseller, Reg. No. 38,318; John C. Stattler, Reg. No. 36,285; Edwin H. Taylor, Reg. No. 25,129; Lester J. Vincent, Reg. No. 31,460; Ben J. Yorks, Reg. No. 33,609; and Norman Zafman, Reg. No. 26,250; my attorneys; and William D. Davis, Reg. No. 38,428; Gary B. Goates, Reg. No. 35,159; Soyeon P. Laub, Reg. No. P-39,266; Thomas X. Li, Reg. No. 37,079; and Edwin A. Sloane, Reg. No. 34,728; my patent agents, with offices located at 12400 Wilshire Boulevard, 7<sup>th</sup> Floor, Los Angeles, California 90025, telephone (310) 207-3800, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected herewith.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

1-00 Full Name of Sole/First Inventor: DELEAM David

Inventor's Signature: Deleam

Date: 11 JAN. 2002

Residence: PERROS GUIREC / FRANCE JRX  
(City, State)

Citizenship: FR  
(Country)

Post Office Address: 11, Rue du Maréchal Leclerc 22700 PERROS GUIREC / FRANCE

2-00 Full Name of Second/Joint Inventor: KOVESI Balazs

Inventor's Signature: Kovesi

Date: 11 JAN. 2002

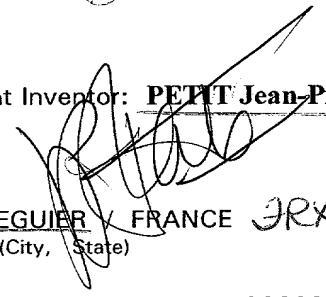
Residence: LANNION / FRANCE JRX  
(City, State)

Citizenship: HU  
(Country)

Post Office Address: 12 Résidence Corlay 22300 LANNION / FRANCE

300

Full Name of Third/Joint Inventor: PETIT Jean-Pierre

Inventor's Signature: 

Date: 11 JAN 2002

Residence: MINIHY TREGUIER / FRANCE *JRX*  
(City, State)

Citizenship: FR  
(Country)

Post Office Address: 10, Cité Zant Erwan 22220 MINIHY TREGUIER / FRANCE

Full Name of Fourth/Joint Inventor:

Inventor's Signature:

Date:

Residence: \_\_\_\_\_  
(City, State)

Citizenship: \_\_\_\_\_  
(Country)

Post Office Address:

10019550.032907  
206220.05567007